TapWristband: A Wearable Keypad System Based on Wrist Vibration Sensing

Jialiang Yan, Siyao Cheng, Yang Zhao, *Senior Member, IEEE*, and Jie Liu, *Fellow, IEEE*

***Abstract*—Fine-grained human motion detection has become increasingly important with the growing popularity of human computer interaction (HCI). However, traditional gesture-based HCI systems often require the design of new operation modes rather than conforming to user habits, thus increasing system learning costs. In this paper, we present TapWristband, a novel wearable sensor-based vibration sensing system that detects finger tapping by measuring wrist vibrations. We first perform real-world experiments to collect measurements for modeling the effects of the tapping motion on wearable wristband sensors including piezo-electric transducer (PZT) and inertial measurement unit (IMU). We find that a damped vibration model can be used to represent the relaxing phase of a vibration response due to tapping motion. Thus, we propose a mutual cross-correlation-based event segmentation algorithm to extract the vibration signal during the relaxing phase. After that, we develop feature extraction and classification algorithms to recognize the tapping patterns of five fingers across twelve key locations of a keypad system. Finally, we performed extensive experiments with thirteen participants to evaluate our system. Experimental results show that our low-cost vibration sensing system can achieve an average accuracy of over 93% with a tapping speed of over 100 taps per minute in real-world tapping scenarios.**

***Index Terms*—Human-cyber-physical systems, tap recognition, wearable sensors, wireless sensing.**

I. Introduction

With the development of Internet of Things (IoT) systems and signal-processing technologies, wearable devices are gaining momentum in various human-computer interaction (HCI) applications [2], [3], [4]. There is a pressing need for compact, readily accessible input devices that take advantage of the fast-developing wearable sensing and computing techniques [1].

Given the pivotal role of motion sensing in HCI, motion detection has been an active research area of HCI in recent years [26], [30], [34]. For HCI technologies based on motion sensing, the ability to detect subtle finger movements is of paramount importance as it directly impacts the overall quality of service provided by the HCI system [6]. Thus, a large amount of research has been undertaken in the field of finger motion detection. For instance, studies in [5] and [6] have proposed to use of sonar and radar sensors for tracking finger movements, specifically for fine motion of finger recognition purposes. While these studies are capable of accurate finger motion tracking, they all use wireless sensors and are sensitive to external disturbances, resulting in unstable performance in real-world HCI scenarios. Recognizing that finger tapping is one of the most effective input methods in HCI, [9] and [14] developed finger tapping detection systems that operate based on the surface vibration mechanism. However, these systems require the placement of sensors at stationary positions on a surface, which compromises the mobility of the systems.

To improve mobility and sensing precision, recent research studies have proposed to use of sensors on wearable devices to detect and recognize finger tapping. For example, in research [8], [13], [16], one hand of a person was used as a keyboard, with the other hand used for typing, and tapping data was gathered by analyzing the vibrations of the hand serving as the keyboard. As another example, a typing ring was developed in [11] to detect human typing motion. Even though these wearable sensor-based systems can improve mobility, they are not consistent with the typing habits of users and thus have not been used as effective HCI input methods. In this work, we aim to develop a vibration sensing-based tap recognition method, which follows the typing behavior of users without adding much learning cost, leading to a more natural and efficient HCI input system.

Given the prevalence of smart devices such as bracelets and watches worn on the wrist of a person, sensors on these wearable devices can be utilized to gather vibration data for tap recognition. When a finger taps on a rigid surface, not only the surface vibrates [15], but also the skin of the person, including his or her wrist and hand, as some of the vibration signals are reflected back [16]. As shown in Fig. 1, tapping fingers generate reflected vibrations in various parts of the hand, which are then propagated to the wrist via different paths [10]. Considering that different tapping vibrations traverse through varying

Received 5 January 2024; revised 17 June 2024; accepted 14 November 2024. Date of publication 20 November 2024; date of current version 6 March 2025. This work was supported in part by the Key Research and Development Program of Heilongjiang Province under Grant 2022ZX01A22, in part by the National Science and Technology Major Project of China under Grant 2021ZD0110901, in part by the National Natural Science Foundation of China under Grant 61972114 and Grant 62106061, in part by the collaborative innovation and promotion system of the modern agricultural industry technology for watermelon and melon in Heilongjiang Province, in part by the Open Competition Project of Heilongjiang Province(China) under Grant 2021ZXJ05A03, and in part by the National Natural Science Foundation of Heilongjiang Province under Grant YQ2019F007. Recommended for acceptance by J. Xu. *(Corresponding author: Siyao Cheng.)*

Jialiang Yan and Siyao Cheng are with the Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China, and also with the National Key Laboratory of Smart Farm Technologies and Systems, Harbin 150001, China (e-mail: yanjialiang029@163.com; csy@hit.edu.cn).

Yang Zhao is with the National Key Laboratory of Smart Farm Technologies and Systems, Harbin 150001, China (e-mail: yang.zhao@hit.edu.cn).

Jie Liu is with the International Research Institute for Artificial Intelligence, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: jieliu@hit.edu.cn).

Digital Object Identifier 10.1109/TMC.2024.3503417

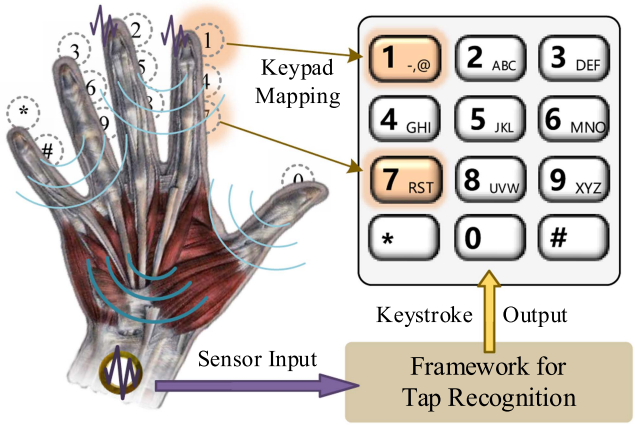

Fig. 1. Overview of vibration sensing-based HCI system: wrist vibration sensing is used to enable mapping of finger tapping on different locations to different keys of a keypad.

distributions of bones and muscles, we ask ourselves a question: is it feasible to differentiate the vibration signals reflected back to the user when tapping with different fingers? Inspired by this question, we propose TapWristband, an innovative system that senses vibrations at the wrist of a person to detect and recognize the finger tapping of the person. More specifically, as illustrated in Fig. 1, sensors located on wrists capture the vibrations caused by finger tapping and taps at different locations are mapped to corresponding keys on keypads. Therefore, TapWristband can be used as a keypad that allows typing on any surface that causes muscle vibrations.

The primary hurdle in achieving a robust and accurate wrist vibration sensing system lies in the stochastic nature of the vibration signal due to tapping. Each individual person has their own tapping habit, which involves varying speeds and strengths. Thus, it is a big challenge to eliminate the random interference in tapping and recognize different tapping motions. Our experiments revealed that the vibration signal from any complex tapping motion can be separated into two phases: a forced vibration phase and a relaxing vibration phase. During the forced vibration phase, the vibration signal has high uncertainty due to different tap intensities, angles, and speeds of tapping. In contrast, the relaxing vibration phase is a process, in which the vibration energy is dissipated across the muscles of the finger, the hand, and the wrist of a person. During this phase, the vibration signal is mainly determined by the initial state of vibration and the unique shape and structure of the hand of the person, making it less prone to other interference. Therefore, relaxing vibration signals are more robust and generalizable than forced vibration signals, in terms of tapping recognition. This leads to two technical issues we aim to address: (1) How to extract the relaxing vibration signal from a tapping event. (2) How to tailor this approach to accommodate the habits of various users.

To deal with these issues, we develop a hybrid damped vibration model for the relaxing vibration phase and propose a wrist vibration recognition framework using the vibration signal in the separated relaxing vibration phase. The framework initiates by identifying event occurrences and mitigating interference through an energy window extraction algorithm and signal decay fitting. Following this, the framework employs a mutual correlation-based exact segmentation algorithm for the adaptive precise extraction of relaxing vibrations. After that, we propose a feature classification based on vibration propagation dispersion and use a feature selection algorithm and joint classification algorithm to recognize wrist vibration caused by finger tapping. Finally, we constructed our system TapWristband for examination based on the above framework using IMU (Inertial Measurement Unit) and PZT (Piezoelectric) sensors. The evaluation of the system involves 13 volunteers and the evaluation shows an accuracy of over 93% and a tapping speed of over 100 keystrokes per minute in real-world typing scenarios.

Our framework is designed on the basis of a wrist vibration model. Our design aims to operate independently of the sensors based on the spring models and not be limited by the type of vibration sensors used in actual HCI systems. To accomplish this, we meticulously selected vibration sensors for our experiments, taking into account several key attributes: widespread usage, market availability, ease of integration, and diverse sensing principles.

In our real-world experiments, we employed two types of vibration sensors: PZT and IMU. PZT sensors offer the advantages of sensitivity, low power consumption, and they can operate passively. However, they have a drawback in that the vibration signals are sensitive to the shape of the sensors. Moreover, pyroelectric effects may interfere with the measurement of mechanical quantities. On the other hand, IMU sensors can be tightly integrated with equipment. They are less affected by changes in sensor shape and are easy to attach to the wrist. However, IMUs do have the drifting issue. Considering the varying vibration sensing characteristics of different sensors, we proposed using sensor data normalization method. This ensures that the signal input to our wrist vibration recognition framework represents a standardized vibration signal. Our validation experiments, which involved constructing separate systems using both PZT and IMU sensors, demonstrated that our framework is effective for sensors based on the spring oscillator principle.

Furthermore, since our framework only senses wrist vibrations, we do not need to ask users to rest their wrists at fixed positions. As long as the height of the wrist relative to the tapping surface is fixed, we also do not require the tapping to start from the same initial positions. The core of the recognition depends on the relative position of the fingers and the palm of the hand. As a result, the TapWristband recognizes flexible tapping and even tapping without visualization on any desktop surface.

In summary, the contributions of this paper are as follows:

- We propose and validate a damped vibration model by performing experiments with multiple vibration sensors to study the effects of tapping motion on wearable wristband sensors.
- We develop a mutual cross-correlation-based event segmentation algorithm to extract vibration signals during a relaxing phase, which is more robust to interference.
- We design and implement a wrist vibration recognition system TapWristband with feature extraction and

classification algorithms to recognize the tapping patterns of five fingers across twelve keys of a keypad system.
- Our extensive experiments and evaluation show that TapWristband system is accurate and convenient to use in real-world scenarios.

The rest of the paper is organized as follows. Section II discusses the related works. Section III introduces the wrist vibration model and model validation. Section IV introduces the overview of the whole system. Section V presents the detailed design of TapWristband. Sections VI and VII show the experimental implementation and the experimental results, respectively. Section VIII discusses the future work. Section IX concludes the whole paper.

II. Related Work

Providing efficient and concise means of human-computer interaction for IoT devices has been a research problem for many years [1]. Traditionally, speech-based methods have been used for interaction [17], but they can lead to serious information leakage and speech input is more sensitive to noise [18]. Thus a variety of gesture-based recognition techniques have been proposed for human-computer interaction [19], [21], [23], [31], [35]. Since the pattern of actions determines the input efficiency, we classify them into two categories based on the difference in the pattern of gesture actions: customized gestures, and keystroke recognition.

A. Customized Gesture Recognition

Initially, gesture sensing techniques were investigated, using a variety of different sensors that were able to recognize several specific gestures [27], [30], [35]. However, these gestures were limited in their ability to convey complex information. Later, handwriting tracking techniques were investigated [24], [25], but the computing and hardware costs were too high to ensure accuracy. In addition, all of the above gesture recognition techniques need to increase the learning cost of the user. Under this thinking, some finger movement tracking algorithms have been studied using various wireless signals, such as frequency modulated continuous wave (FMCW) [33], millimeter wave [6], [28], and ultrasound [5], [34], to perform high-precision heuristic hand gesture recognition. These methods are efficient in detecting handwriting movements, but require expensive equipment, higher computing costs, and can only be applied to some special occasions.

B. Keystroke Recognition

Unlike customized gesture recognition, keystroke recognition does not require additional learning costs for the user. Therefore, it is more user-friendly and has high input efficiency. Initially, algorithms for detecting tapping movements by surface vibration were provided [10], [14], [29], [36], [37]. However, these techniques resulted in a keyboard that was difficult to move around and was susceptible to desktop noise, resulting in a poor user experience. So some people study the keystroke recognition technology of wearable devices [19], [22], [38], [39], but these wearable devices need to be specially customized and do not meet people's daily habits, which increases the cost of wearing and using the user. Then someone thought of using a part of the body as a keyboard, through the sensor to collect body information to recognize the keys, these keyboards have the use of fingers, the back of the hand has been arm [13], [16], [26], [41] and so on. They do solve the problem of carrying the keyboard, but using the body as a keyboard does not work stably for a long time and is not a comfortable option.

TABLE I
Motivation of the TapWristband

Items	Efficiency	Comfort	Usability
Heuristic Gesture [6], [28]	Low	Yes	Medium
Surface Vibration [29], [36]	High	Yes	Low
Wearable Keystroke [13], [26]	High	No	High
TapWristband	High	Yes	High

Inspired by the aforementioned studies, we found the closer the gesture patterns are to those of a real keyboard, the more confident and motivated the user is, and thus the better it works. As shown in Table I, TapWristband aims to mimic the standard gestures of a one-handed operation of a mini-keyboard by identifying tap vibrations from the wrist. This technology can be operated on any hard surface or objects, and even without visual observation. The excellent integrability of TapWristband allows users to interact with it as if they were typing in a traditional way, thus enhancing the usability of this technology and unlocking its potential.

III. Vibration Modeling and Validation

In this section, we formulated a model to describe the vibrations generated by finger tapping at the wrist. Subsequently, we validated our constructed model across multiple dimensions.

A. Vibration Mode

We have observed that the vibration produced by a person tapping down with a finger varies depending on the finger and the part of the finger used (fingernail, fingertip, fingertip belly), and this difference is transmitted to the wrist. In order to distinguish between vibration signals resulting from taps made by different fingers and at different locations, it is necessary to construct a model that accurately represents the vibration produced during the tapping process. The entire tapping action includes the finger forcefully moving upward, then moving downward, and finally the fingers contact the desktop to make a tap. As shown in Fig. 2, the vibration signal can be divided into two segments: the first being the muscle-driven movement of the hand, which we call the forced vibration, and the second being the relaxing vibration of various parts of the hand, which we call the relaxing vibration. Throughout the tapping event, the wrist experiences pulling forces from the entire mass system of the hand [22]. Consequently, the coupled oscillations at the wrist can be utilized to analyze the finger's state [32].

Observations indicate that the relaxing vibration possesses objective properties with enhanced generalization capabilities

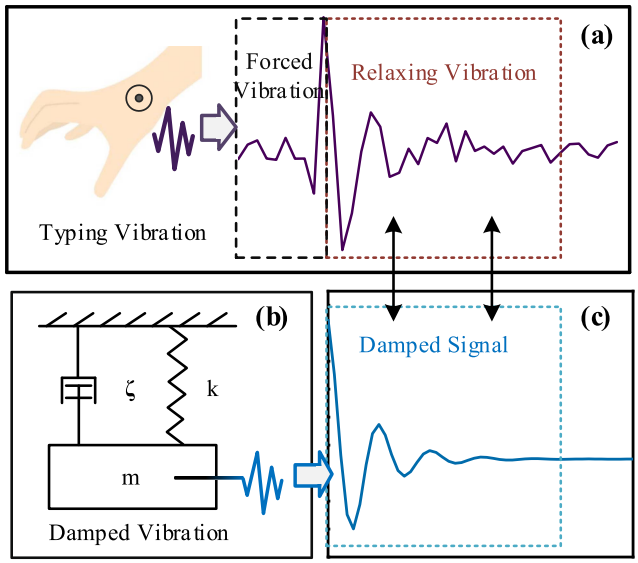

Fig. 2. Damped vibration model of hand vibration signals. (a) The vibration signals of the tapping wrist, where the signal in the black box indicates the forced vibration signal and the signal in the brown box indicates the relaxing vibration signal; (b) Damped vibration model of a single mass; (c) Results of fitting multiple damped vibrations of a wrist relaxing vibration.

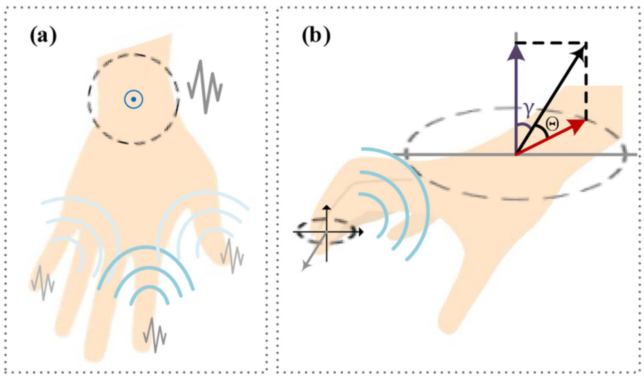

Fig. 3. Schematic diagram of generation and propagation of hand damped vibrations. (a)The vibration at the wrist is affected by the vibration and propagation of the mass in the hand; (b)Vibration is concentrated in the z-axis component at the wrist.

and stability, in contrast to the initial muscle-driven phase. However, the complexity of the human hand, which consists of 27 bones and corresponding soft-tissue structures, complicates the direct modeling of hand vibration equations. We use common market sensors such as IMU and PZT vibration sensors, which are based on the spring oscillator principle, as shown in Fig. 2. In the relaxing vibration, there are no more forcing factors. The initial conditions are determined by the initial potential energy. Therefore, we can consider the hand vibration as a composite damped vibration system. Finally, the wrist vibration signal obtained by the sensor is the combination of Z-axis components of multiple damped vibrations. As shown in Fig. 3, the vibration at the wrist is the projection of the signals from various parts of the hand in the z-axis direction. Therefore, the coupled vibration at the wrist can be regarded as the synthesis of multiple damped vibrations in the same direction, which can be expressed as:

$$x(t) = \sum_{i=1}^{n} x_i(t)\cos(\gamma_i) \tag{1}$$

where $x(t)$ is the displacement of the object at time t, γ_i denotes the angle between the vibration of the ith mass and the z-axis, and $x_i(t)$ denotes the damped vibration of the ith mass. The sensor we use is the principle of spring oscillator, thus $x_i(t)$ can be expressed as [12]:

$$x_i(t) = A_i e^{-\zeta_i \omega_{n_i} t}\cos\left(\omega_{d_i} t - \phi_i\right) \tag{2}$$

where A_i and ϕ_i correspond to the amplitude and initial phase of the ith signal, respectively. The damping ratio is represented by ζ_i. The undamped natural frequency is denoted by ω_{n_i}, while the damped frequency is represented by ω_{d_i}. Note that in the context of the ith damped oscillation, ζ_i represents the proportion of the damping force to the critical damping force. The system transitions into an overdamped state and halts oscillation when $\zeta_i > 1$. At $\zeta_i = 1$, the system achieves a critically damped state, eliminating oscillation. The system is deemed underdamped and displays oscillatory behavior when $\zeta_i < 1$, which is also the case with this system. The undamped natural frequency, denoted by $\omega_{n_i} = \sqrt{\frac{k_i}{m_i}}$, dictates the oscillation frequency of the particle in the absence of damping, where m_i signifies the mass of the particle and k_i is the spring constant. The damped frequency, represented by $\omega_{d_i} = \omega_{n_i}\sqrt{1-\zeta_i^2}$, determines the oscillation frequency of the system in the presence of damping. An interesting observation from this study is that post the finger's strike and subsequent relaxing vibrations, the vibration parameters of the hand mass system, barring the amplitude A_i and initial phase ϕ_i, are self-determined and remain unaffected by muscle movement.

The coupled signals have two main sources: the first source is the damped vibrations generated by the tissues at the wrist, and the other one originates from the propagation of vibrations from other fingers of the hand. The information on the intrinsic frequency and damping of the different signals is included in the corresponding frequencies, which can be visualized by the Fourier transform. The coupled vibration signal after a Fourier transform can be represented as:

$$X(\omega) = \int_{-\infty}^{\infty} \sum_{i=1}^{n} \left[Z_i e^{-\zeta\omega_n t}\cos(\omega_d t - \phi)\right] e^{-i\omega t} dt \tag{3}$$

where $X(\omega)$ denotes the Fourier transform of the damped oscillation, with ω representing the frequency. Eq. (3) provides a representation of the damped oscillation $x(t)$ in the frequency domain, effectively describing the distribution of the damped oscillation across various frequencies.

It is worth mentioning that the information such as initial phase ϕ, intrinsic frequency ω_n, and damping ζ of the vibration signal is closely related to the soft tissues and mass, etc., in which the signal is generated. Moreover, during propagation, vibrations have different speeds, attenuation, and frequency characteristics as they propagate through bone, skin, soft tissue, etc. Thus, different ω can reflect the characteristics of the corresponding vibration signal in its frequency band range [42].

Different keystrokes excite different vibration generations and pass through different tissues, and thus different keystrokes can be distinguished. In addition, the muscle, skin, and bone tissues of different human hands are similar in structure, although slightly different, and thus the relaxing vibration signals have a strong generalization ability.

B. Model Validation

In this subsection, we conduct experimental validation of the aforementioned models and analysis. Specifically, we explore the stability, distinguishability, and generalizability of finger tapping events and their associated coupled vibration signals.

1) Model of Finger Tapping Events: First, we need to define specific tapping gestures that can be applied within our sensing framework. These gestures need to have the following two key properties. (1) The set of gestures can be differentiated by wrist tapping; (2) These gestures align with the user's habitual behavior. Our observations and research indicate a standard fingering pattern when interacting with a keypad. Specifically, the index finger is used for keys 1, 4, and 7, the middle finger for keys 2, 5, and 8, the ring finger for keys 3, 6, and 9, and the thumb and little finger are used for function keys based on habit.

This fingering pattern aligns well with our requirements, necessitating only minor adjustments for applicability. Our gesture assumes a fixed wrist and keyboard position as depicted in Fig. 1. Keys 1, 2, 3, and 10 are struck with the index finger, middle finger, ring finger, and the pad of the little finger, respectively. Keys 4, 5, 6, and 11 are struck with the index finger, middle finger, ring finger, and the tip of the little finger, respectively. Keys 7, 8, and 9 are struck using the nails and fingertips of the index, middle, and ring fingers, and the 0 key is struck using the thumb.

It's worth noting that in this method, we do not need to ask users to rest their wrists at fixed positions. As long as the height of the wrist relative to the tapping surface is fixed, we also do not require the tapping to start from the same initial positions. The key requirement is that the volunteers should maintain a fixed tapping posture. Unless otherwise stated, the stability, distinguishability, and versatility of the method's signals are subsequently examined in depth according to this general scheme.

2) Stability: First, we ask ourselves if different finger tap motions cause distinguishable and stable vibration signals. To ascertain the stability of the vibration signal, we conducted the following experiment. A volunteer was asked to tap the same spot with the same finger five consecutive times, with a few seconds intervals between each tap. The results were superimposed as shown in Fig. 4. It was observed that the forced vibration segment of the signal seemed unstable. However, the relaxing vibration segment produced a stable signal when the same tapping posture was maintained. This stability can be attributed to this part of the vibration being less susceptible to muscle interference.

3) Distinguishability: To confirm that finger tapping at various locations generates distinguishable differential vibration signals at the wrist, we conducted an experiment. A volunteer

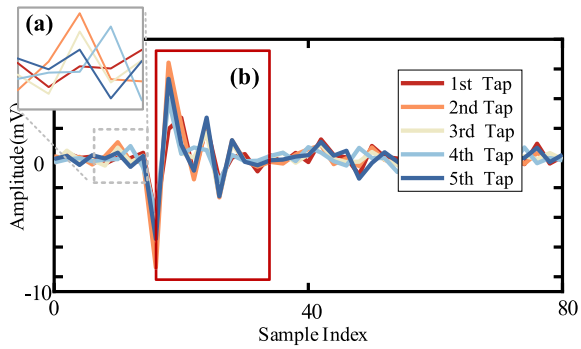

Fig. 4. Vibration signals from multiple taps at a uniform location. (a) Forced vibration signal; (b) Relaxing vibration signal.

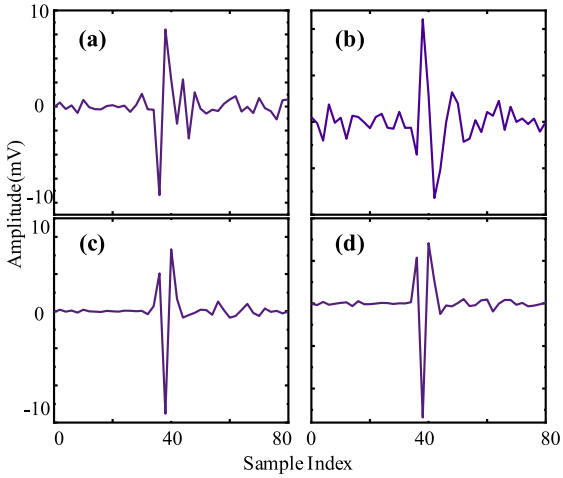

Fig. 5. Different finger tapping signals. (a) Tap key 4 with the index finger; (b) Tap key 8 with the middle finger; (c) Pan hand and tap key 4 with the middle finger; (d) Tap key 5 with the middle finger.

was instructed to execute four consecutive taps at different positions on a table. The signals were then captured, and the results are presented in Fig. 5. Fig. 5(a), (b) illustrates the variance in signals from different fingers tapping at distinct locations. A review of Fig. 5(a)–(c) reveals that the tapping signals of different fingers at the same position are distinguishable. Given that the vibration signals are influenced by the relative positions of the fingers and the hand, the tapping of the same finger at different positions also varies, as evidenced by the differences in Fig. 5(b)–(d). Notably, Fig. 5(c), (d) display the results of a single finger tapping at different positions, yet the signals exhibit a strong resemblance. This underscores the fact that the variation in the vibration signal is dictated by the relative positions of the fingers and the wrist. In this experiment, there is no stipulation for the wrist position to remain constant, and individuals will tap different points within a close range using comparable postures. It is this similarity in posture that results in the resemblance between the signals from the two taps.

4) Generalizability: We divide the signal into two parts, relaxing vibration and forced vibration, and in this experiment, we need to verify whether increasing the forced vibration signal will cause the universality of the signal to deteriorate. In this

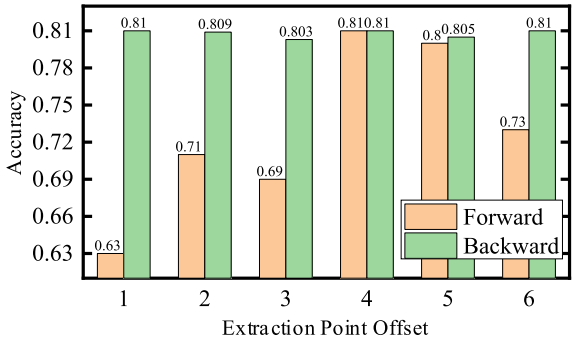

Fig. 6. Effect of the coordinates of the forward and backward movement of the segmented signal on the accuracy of recognition.

experiment, we recruited three volunteers, asked them to rest their wrists in the designated area, and then tap 10 times at each position on the keyboard as shown in Fig. 1. Then, we collected their vibration signals when they tapped at different key positions. In order to control the variables, we simultaneously reduced the signal at the end of the relaxing vibration while increasing the forced vibration signal before the relaxing vibration. In this way, we can ensure that the length of the signal is the same, that is, the amount of information is the same. We perform high-pass filtering on these signals and roughly align them using cross-correlation [40]. Subsequently, we truncate the signal and extract its amplitude spectral density (ASD) as a feature value, and then classify it using a random forest classifier [43]. The average accuracy of 5-fold cross-validation is shown in Fig. 6.

Initially, we fixed the post-signal interception point to confine the interception range, with the results indicated as yellow bars in Fig. 6. It can be observed that the classification effect is suboptimal at the outset when the signal contains forced vibration. However, as the interception point shifts, the forced vibration diminishes and eventually vanishes, leading to the best classification result. As the interception point continues to shift, the classification results are optimized. However, as the effective information (relaxing vibrations) decreases, the classification results deteriorate.

To eliminate the interference of interception length, we fix the former segmentation point at the optimal segmentation position and then use the corresponding size of the interception length to conduct the experiment, with the results displayed in Fig. 6. At this stage, the classification results are not significantly affected, which further validates that the forced vibration in the first half of the segment has poor generalizability to the signal, while the relaxing vibration exhibits strong generalizability.

IV. TapWristband System Overview

Briefly, we introduce TapWristband, as depicted in Fig. 7. This system captures the vibrations produced by finger taps and channels these vibration signals into the tap recognition framework. This framework subsequently translates the finger taps on the table into the output from the 12-key keypad. The tap recognition framework can be divided into five major components as follows:

- *Pre-processing:* The signals acquired by the sensors are first removed from the background noise, after which the signals with different dimensions are converted into a uniform time-domain displacement signal $x(t)$, and finally narrow-band filtered to facilitate subsequent tap detection.
- *Tap Detection:* An energy-based double-threshold sliding window is used for pre-detection, followed by the use of decay fitting to determine the occurrence of taps.
- *Relaxing Vibration Segmentation:* An adaptive correlation function $C(t)$ is generated using the decay fitting results and the selected signal $s(t)$, and a cross-correlation algorithm is used to determine the signal segmentation points.
- *Feature Extraction:* The features of the taps are extracted from both frequency and time dimensions and feature selection is performed on them.
- *Joint Classifier:* Construct a joint classifier to classify and recognize taps by a weighted combination of several classic classifiers.

V. Methodology

In this section, we delve into the characteristics of the coupled vibrations obtained at the wrist. To validate our model, we used two sensor systems - PZT and IMU - to collect vibration data from the wrist. The signals collected by these sensors are represented as $X_{PZT}(t)$ and $A_{IMU}(t)$, respectively. Following the pre-processing of the sensor signal, we obtain a vibration signal $x(t)$ containing coupled vibrations. To capture and dissect the effective portion of this vibration, we have developed a sequence of algorithms.

For recognition of taps, it is necessary to monitor $x(t)$ for occurrences and intercept them. In general, time-window-based energy monitoring is a commonly used method because the vibration at the wrist is a bursty finite energy signal. But there exist several challenges in accurately identifying the tapping event within the $x(t)$ signal:

- The duration of the tapping event is unpredictable: While the timing of the relaxing vibration is generally stable, it frequently interacts with other disturbances, leading to events of varying lengths. These events typically last between 0.02 s and 0.2 s [15].
- Tapping events exhibit a high degree of coupling: The tapping action is inherently a dynamic force process, and our aim is to detect its relaxing vibration component. However, this component is often enveloped by these interconnected signals.
- Effects of passive taps: The signal $x(t)$ will encompass passive vibration signals produced by various situations such as vibrations from a table, collision sensors, and arm collisions, in addition to our active tapping signals. These passive vibration signals bear a high resemblance to our actual signals, making their elimination through an energy windowing method very challenging.

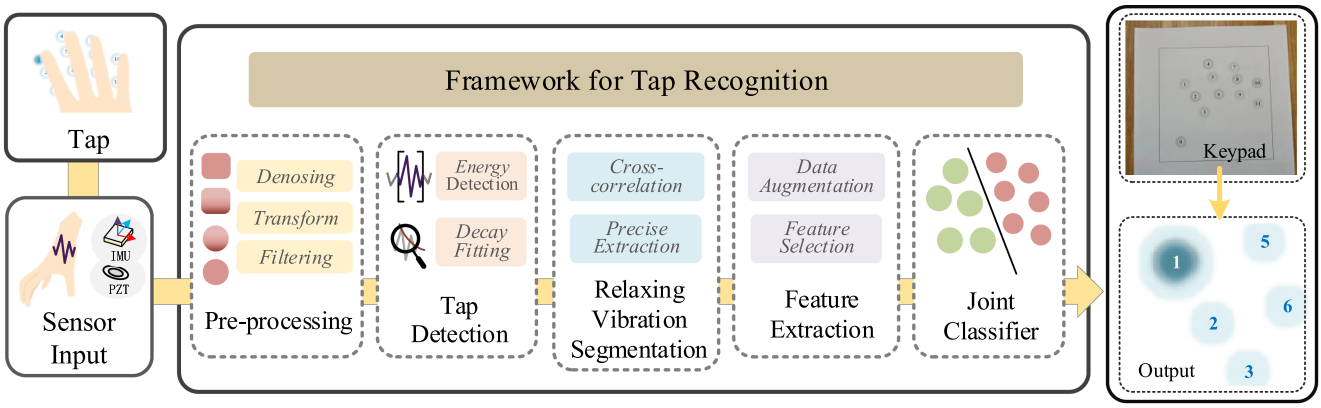

Fig. 7. The Overview of TapWristband framework, which includes five major components: *Pre-processing*, *Tap Detection*, *Relaxing Vibration Segmentation*, *Feature Extraction* and *Joint Classifier*. The image in the upper right corner indicates the keypad used to indicate where to tap.

To tackle the aforementioned challenges, we carry out data pre-processing and standardize the signals from various sensors to ensure that the subsequent tasks are sensor-independent. Following this, we ascertain whether a tapping event has occurred, eliminate the impact of false taps, and then execute decoupling to extract valuable information. This approach offers two benefits: first, breaking down the problem simplifies each component and enhances the extraction success rate; second, it guarantees the system's real-time performance. We have devised the entire segmentation process, as illustrated in Fig. 8, to overcome these challenges and to acquire the data of isolated relaxing vibrations of the wrist.

A. Pre-Processing

1) Denosing: Since vibration sensors are not sensitive to high-frequency signals such as ambient noise, our signals are mainly disrupted by low-frequency noise originating from actions such as body movements and clothing friction. Fortunately, such noise is generally below 10 Hz [37], and according to the a priori information obtained from the tests, our vibration signal is predominantly within the $100-300$ Hz range. Therefore, for the data $X_{PZT}(t)$ and $A_{IMU}(t)$ acquired by the sensors, we use a high pass filter with a cutoff frequency of 20 Hz to eliminate the random noise introduced by the environment.

2) Transform: The operating principle of PZT is based on the piezoelectric effect. When vibration occurs, the PZT sheet undergoes deformation, generating a signal $X_{PZT}(t)$ that signifies the displacement $x(t)$. The signal $A_{IMU}(t)$, obtained by the IMU during vibration, denotes the acceleration $a(t)$. Due to the IMU and PZT vibration sensor principles are based on the spring oscillator principle, $a(t)$ can be transposed to displacement using the equation:

$$x(t) = \int \left(\int a(t)dt + v_0 \right) dt + s_0$$

where $x(t)$ represents the displacement at time t, $a(t) = A^o_{IMU}(t)$ signifies the acceleration at time t, v_0 is the initial velocity, and s_0 is the initial displacement. In this study, we assume both the initial velocity v_0 and initial displacement s_0 to be 0, and the transformed signal is denoted as $X^o_{IMU}(t)$.

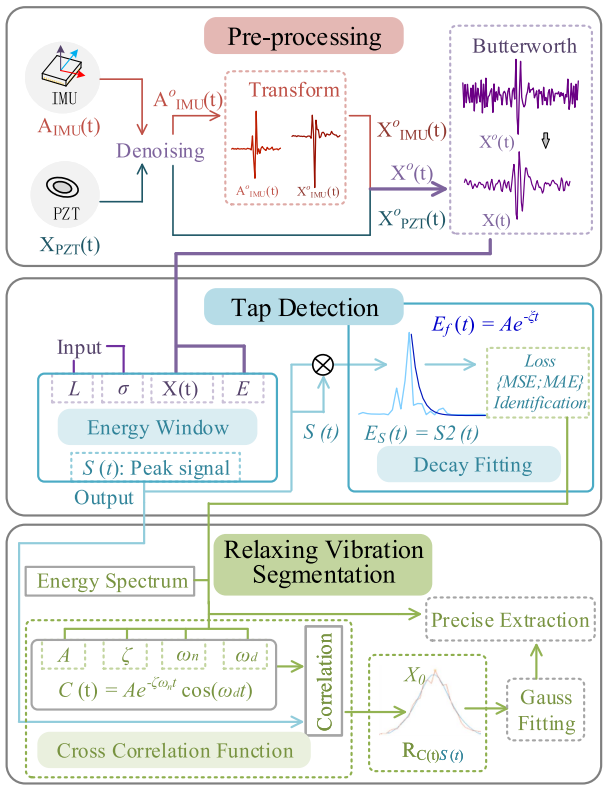

Fig. 8. Framework of Precision Tapping Event Extraction, which consists of three main parts, *Pre-processing* for denoising and unifying the signals from different sensors, *Tap Detection* for monitoring and localizing the occurrence of events, and *Relaxing Vibration Segmentation* for intercepting the valid signals out.

3) Filtering: Since PZT directly obtains the displacement signal thus we used PZT to acquire these typical passive knocking noise signals $X(t)_{PZT}(t)$ as shown in Fig. 9. The sampling rate of the PZT setting is $860Hz$.

According to the description above, the noise generated by sudden shocks to the sensors was the most disturbing for us. During the experiment, we encountered several types of typical

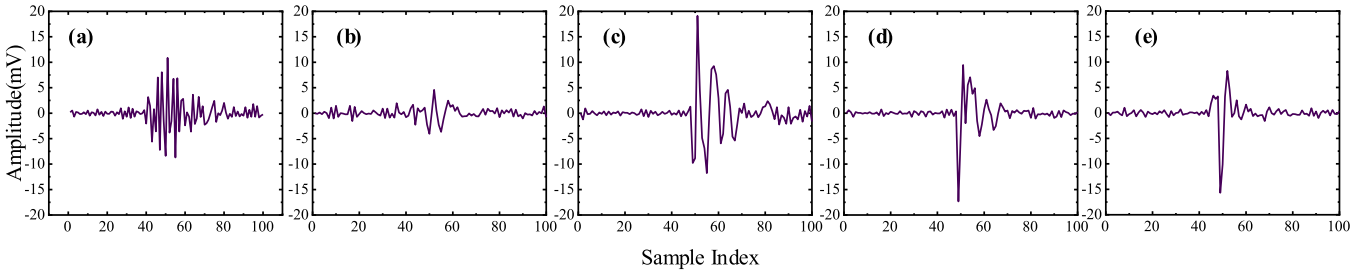

Fig. 9. Distribution of different types of noise in the time domain. (a) Slapping on the arm; (b) Hitting the body; (c) Knocking on the table; (d) Collision directly on the sensor; (e) Normal tap.

noise signals as shown in Fig. 9. These signals were generated by slapping on the volunteer's arm, hitting the volunteer's body by other people or objects, knocking on the desktop, and collision directly on the sensor, respectively. Since PZT directly obtains the displacement signal thus we used PZT to acquire these typical passive knocking noise signals $X(t)_{PZT}(t)$ as shown in Fig. 9. The sampling rate of the PZT setting is $860Hz$. For these signals generated by human vibration and direct impacts on the sensors, there is a significant frequency difference between them due to their different propagation paths. Given the Butterworth filter's advantage of a flat frequency response in the passband, which results in less distortion and minimal signal impact, we employ this filter to eliminate this portion of the burst signal in this experiment. We convert it into a digital bandpass filter using the bilinear transformation. We set its pass and cutoff frequencies to $150Hz$ and $280Hz$, represented by W_c and W_p, respectively. The filtered outcomes of the two sensor signals can be combined and denoted as $X(t)$. Our filter can achieve approximately a $15dB$ improvement in the signal-to-noise ratio. An additional benefit of employing this narrow-band filter is that it eliminates intricate details from the vibrations, thereby facilitating a more precise fitting in subsequent decay fitting.

B. Tap Detection

To identify the occurrence of the tapping event and approximately ascertain its location, we have devised the detection process, as shown in Fig. 8. For complete detection of tapping actions, the algorithm (refer to Algorithm 1) given in this section is based on a sliding window, which contains two modules, respectively. These two modules, *Energy Window* and *Decay Fitting*, are elaborated upon in the subsequent sections.

1) Energy Window: Tapping generates a burst signal, and the energy window serves as an effective event detection method, despite its inability to decouple the free vibration of the tap. Let's assume that L represents the target event width, t is the time of the signal under detection, and $X(t)$ is the signal corresponding to the time t. E_0 and σ are the average energy of the signal and the standard deviation of the signal energy, respectively, obtained a priori. Subsequently, we devise an energy window algorithm based on double pointers, which encompasses the following four primary steps.

First, compute the average energy within the time window $[i, i+L]$, represented by E_w. This can be formulated as follows:

$$E_w = \frac{1}{2}\sum_{t=1}^{L}\frac{X(t)^2}{L}$$

Second, if $E_w > E_0 - r\sigma$, we regard the time of the most recent high-energy occurrence, recorded by $Peak_0$, as the potential starting point for a peak. Here, E_0 and σ represent the mean and variance of the tap signals obtained from the training dataset, respectively. Additionally, r is a constant whose value is determined by the proportion of false taps in the signal, which is typically low, leading to r assuming a value of 2.

Third, given that the energy triggered by a tapping event should not be excessively high, we classify the event as false tapping noise and set $flag\_high$ to '$False$' when $E_w > E_0 + p\sigma$. Here, p is a constant whose value is contingent on the quantity and intensity of high-energy false taps in the signal. To guarantee the detection of as many events as possible by the decay fit, p is typically set to 3.

Finally, if $E_w < E_0 - r/\sigma$ indicates the end of a burst, it is then detected by a decay fitting if the event duration $i - Peak_0$ meets the input conditions and the event's energy is not overly high. The signal that successfully passes the detection is recorded in S. Subsequently, $flag\_high$ and $Peak_0$ are updated to serve as initial values for the forthcoming round of event detection.

2) Decay Fitting: To precisely determine the occurrence of an event, it's essential to detect the relaxing vibration segment. Relaxing vibration is a process where the energy decays from its peak, and false tap signals exhibit a different energy decay pattern compared to real taps due to the differing vibration processes. As shown in Fig. 10 (Fig. 10 in the new manuscript), Fig. 10(a) represents the original signal. This signal needs to be converted into an energy profile and then fitted to its envelope. The fitting result is shown in Fig. 10(b). To identify the energy decay in the coupled signal, we fit the intercepted signal $E_S(t) = s(t)^2$ in segments using the energy decay function, which is responsible for forming the damped vibration. This can be expressed as follows,

$$E_f(t) = Ae^{-\zeta\omega_n t} \tag{4}$$

where A is the amplitude, ζ is the damping ratio, and ω_n is the natural frequency. This formula describes the decay of energy in a damped oscillation over time.

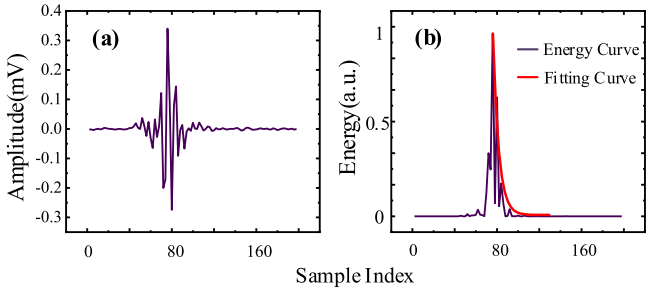

Fig. 10. Raw signal after filtering and fading fitting. (a) Raw signal; (b) Fading fitting result.

Algorithm 1: Tap Detection Algorithm.

Input: $X(t)$: Time domain signal after denoising
L: length of windows. E: average energy
σ: standard deviation of energy
Output: S: Set of signal sigments that contains the tap event

$Peak_0 \leftarrow 0$; $flag\_high \leftarrow True$; $S \leftarrow \{\}$;
for i in t **do**
  calculate the energy E_w of $[i, i+L]$
  if $E_w > E_0 - r\sigma$ **then**
    $Peak_0 \leftarrow i$
    if $E_w > E_0 + p\sigma$ **then**
      $flag\_high \leftarrow False$
    end if
  else
    if $L < (i - Peak_0) < 2L \&\& flag\_high$ **then**
      $s(t) \leftarrow X[Peak_0 : i + L]$
      Fading Fit Detection for $s(t)$
      Add event $s(t)$ to collection S
    end if
    $flag\_high \leftarrow True$
    $Peak_0 \leftarrow i$
  end if
end for
return R

It's noteworthy that the coupled vibration is composed of two parts: the local vibration and the other is the vibration by propagation. While the vibrations transmitted through different tissues convey crucial structural information, the local vibrations are significantly more energetic than the propagated vibrations in terms of energy. This allows us to utilize this more stable segment for fitting to identify the occurrence of a tapping event, which is why we opted for a filter with a narrower passband in the preceding filtering phase. We employ the mean square error (MSE) and the mean absolute error (MAE) as a loss function to measure the discrepancy between the attenuation of the vibration signal and the anticipated value, which can be articulated as follows:

$$Loss\_MSE = \frac{1}{n}\sum_{i=1}^{n}\left(X(i)^2 - E_f(i)\right)^2$$

$$Loss\_MAE = \frac{1}{n}\sum_{i=1}^{n}\left|X(i)^2 - E_f(i)\right|$$

where n represents the total number of samples, and $E_f(i)$ is the value obtained from the fitting process. The MSE is highly sensitive to errors, while the MAE is more susceptible to outliers. During our analysis, we fit the segments of $s(t)$ using σ_MSE and σ_MAE as thresholds for MSE and MAE, respectively. If any segment exhibits an MSE or MAE below these thresholds, we interpret this as the occurrence of a tap event in $s(t)$. To ensure accuracy, we perform a stepwise fitting over the signal $s(t)$, maintaining a constant length. We then fit the interval that yields the minimum value of $Loss = Loss\_MSE + Loss\_MAE$, which corresponds to the position of the tap event.

C. Relaxing Vibration Segmentation

All signals recognized as tapping events are captured and included in the set S. The relaxing vibration signal, which is essential for these events, is often mixed with various disturbances and needs to be isolated.

Interestingly, during the relaxing vibration time, there is no interference from other hand vibrations. This means that the interfering signals within the relaxing vibration signal can be distinguished on the time axis. This observation is significant as it implies that our decoupling process can be viewed as a more refined segmentation, specifically, identifying the start of the relaxing vibration signal. To localize within a signal, signal correlation is a commonly used method [40]. In the subsequent sections, we will elaborate on the cross-correlation functions and methods developed in this study.

Cross-correlating Function: Cross-correlation is a measure of similarity between two signals as a function of the displacement of one relative to the other. The equation for calculating the cross-correlation of two discrete signals, $x(t)$ and $C(t)$, can be represented as follows,

$$R_{xC}(t) = \sum_{n=-\infty}^{+\infty} x(t) \cdot C^*(n-t) \tag{5}$$

where $x(t)$ and $C(t)$ represent two discrete-time signals. $C^*(n-t)$ is the complex conjugate of $C(t)$, which is delayed by k time units. $R_{xC}(t)$ denotes the cross-correlation result.

The outcome, $R_{xC}(t)$, reflects the similarity between the correlation function and the signal under measurement. In other words, if the fixed portion of the signal under measurement closely resembles the correlation function, we can pinpoint the exact location of the relaxing signal via the peak of $R_{xC}(t)$, thereby extracting the necessary information. However, we cannot predict the tap signal to establish the correlation function of the response.

Fortunately, the coupled signal we require is made up of two components: the localized relaxing vibration and the propagating relaxing vibration. The localized relaxing vibration takes place first and is precisely located at the termination of the forced motion. Due to the high energy of the localized relaxing vibrations, the impact of the propagating relaxing vibrations can be overlooked in the cross-correlation. As a result, we can

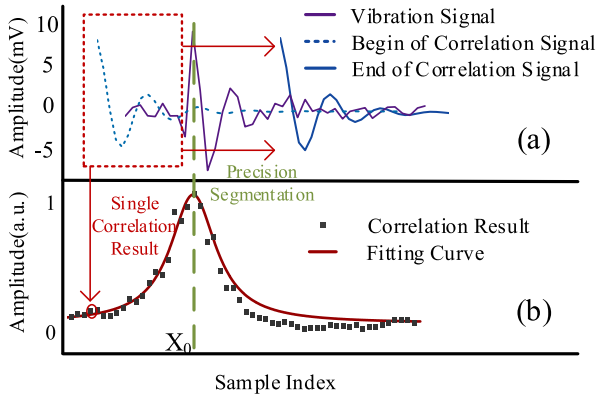

Fig. 11. Cross-correlation results and Gaussian fitting results, where the green dotted line represents the location of the identified segmentation.

pinpoint the location of the signal segmentation by selecting the correlation function solely for the localized relaxing vibrations.

In the preceding section, we used decay to fit the local vibrations. The outcome of this fitting process can be utilized to reconstruct a correlation function for cross-correlation. According the fitting (4), we derived A and $\zeta\omega_n$. To individually obtain the values of ζ and ω_n, we input the signal $s(t)$ into (3) to acquire its energy spectrum. The highest part of the energy spectrum is deemed to be the frequency of the local signal, which is ω_n. These parameters are then input into (1) to reconstruct the local signal, denoted as $C(t)$. This is also our cross-correlation signal. It's important to note that the reconstructed local signal is not dependent on volunteers. The reconstructed signal is achieved by fitting a correlation function that can be adaptively adjusted for different users.

Precise Extraction: To optimize the correlation between the signal $s(t)$ and the correlation function $C(t)$, we configure the filter based on the frequency range of the cross-correlation function $C(t)$ and apply it to the signal to be measured, $s(t)$. The result of this filtering process is denoted as $s_{MF}(t)$. By introducing this signal $s_{MF}(t)$ and the correlation function $C(t)$ into (5), we obtain the cross-correlation spectrum of the signal, as shown in Fig. 11. The trajectory of the correlation function $C(t)$ is represented by the blue dashed line transitioning to the blue solid line in Fig. 11(a), and the correlation result at each position is the vertical coordinate of the corresponding point in Fig. 11(b). Given that the fluctuating noise of the signal adheres to a normal distribution, we fit a Gaussian to the result, which is represented by the red line in Fig. 11(b). The horizontal coordinate at the peak of the fitting result, X_0, is considered the location of the strongest correlation between the two signals. According to Fig. 11, the split point can be calculated as $X_0 + t_0$. The signal following this split point, which represents the relaxing vibration, is required for tapping position recognition.

D. Feature Extraction

1) Fundamental Feature: Due to the dispersive characteristics of vibration signals, vibrations induced by tapping at various locations may exhibit distinct spectra. The amplitude spectral density (ASD) is typically employed as a fundamental feature of the frequency response for vibration signals [8], which can be directly obtained from segmented signals. However, the arrival time of different frequencies at the sensor can vary [42]. Since ASD cannot depict the temporal distribution of vibrations, we utilize the correlation spectrum (CD) as the primary characteristic of the time-domain feature.

For the segmented vibration signals that have been extracted, the ASD and CD can be readily derived as mentioned above. However, it is important to note that the uniqueness of features can be influenced differently by various frequency bands. Certain frequency points exhibit significant variations, even when the tapping position remains constant. These are referred to as *unstable frequency points* and can hinder the accurate characterization of the tapping position. Conversely, there are *retarded frequency points* that show minimal change when tapped at different locations. For optimal tap recognition, it is crucial to minimize the impact of these influential frequency points as much as possible.

2) Data Augmentation: Despite the transducer's sampling rate encompassing the entire signal interval, the frequency response can be significantly impacted by jump points induced by random noise. This results in an increase in *unstable frequency points* and *retarded frequency points*. To mitigate this effect, it becomes necessary to thicken the data via interpolation. Cubic spline [20] interpolation generates an interpolation function with consecutive first- and second-order derivatives at each data point, which ensures the smoothness of the interpolation curve. This avoids problems that may occur when using a global interpolating function, such as the Lunge phenomenon.

3) Feature Selection: Most feature selection methods focus on finding the smallest optimal subset based on classification accuracy. However, the limited accuracy of a particular model is not sufficient to confirm that a feature is irrelevant. Therefore, we employ the Boruta [45] algorithm to identify all relevant classification features. The algorithm relies on a computationally efficient random forest classifier to iteratively discard less relevant features. After feature selection, the feature vector initially consisted of 402 features, which were later reduced to 162 and fed into our classification model.

E. Joint Classifier

We used the following predictive models, all of which have shown excellent performance in classification recognition:

- Support Vector Machine (SVM): it is a statistical learning method with a linear kernel that determines the optimal hyperplane to classify the categories by maximizing the interval between the nearest points.
- Random Forest (RF): it fits a specific decision tree classifier on subsamples and uses averaging to reduce overfitting. We set the estimate to 200 and use the entropy criterion for prediction.
- Linear Discriminant Analysis (LDA): it creates linear decision boundaries to separate classes by utilizing Bayes' rule and approximating the class-conditional density of the samples. We choose singular value decomposition as the solver.

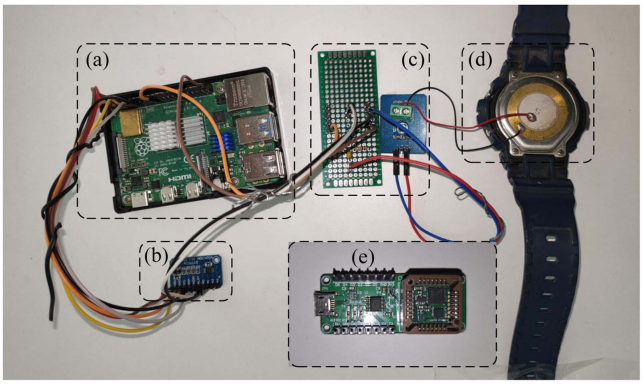

Fig. 12. Platform for experiments: (a) Raspberry Pi 4B; (b) ADS 1115; (c) Sampling circuit; (d) PZT-5H-Single; (e) IMU sensor JY931.

- K-Nearest Neighbors (KNN): it determines the category of a sample by unknown the k nearest neighbors of the sample at a time after mapping the sample into an n-dimensional space.

According to our empirical analysis, RF is best suited for classification, followed by LDA and SVM. Therefore, we assign different weights to different classifiers (SVM, RF, LDA, KNN) to further improve the classification accuracy.

VI. Implementation

In this section, we describe the implementation of TapWristband in terms of system framework, hardware design, and architectural optimization, respectively.

A. System Framework

The framework of our system implementation is divided into two parts, one part is the sensor system as a signal acquisition module and the other part is the processor deploying this tap recognition framework. The tap recognition framework is deployed on the Raspberry Pi 4B platform. The signal acquisition module uses two different sensor schemes to collect data from the wrist. The first scheme uses a PZT with a diameter of 1.5 cm to collect data. The signal from the PZT (PZT-5H-Single) is converted to a digital signal by an analog-to-digital converter (ADS 1115). The digital signal is then amplified four times using a programmable gain amplifier (PGA) integrated into the analog-to-digital converter. Another method is to use an IMU sensor (JY931), which can be directly connected to the Raspberry Pi 4B platform due to its built-in data acquisition circuitry and analog-to-digital conversion function, as shown in Fig. 12. In order to ensure that there is no loss of information, we set the sampling rate of ADS1115 to 860 times per second, and the sampling rate of JY931 to 1000 times per second, according to the sampling theorem. For the sake of portability and mobility, the whole data acquisition is carried out on Raspberry Pi.

B. Hardware Design

Given the properties of the PZT, when the wrist vibrates due to finger movements, the PZT worn on the wrist undergoes deformation due to pressure. Consequently, positive and negative charges emerge on its two opposing surfaces, indicating that the PZT is charging. Conversely, when the pressure dissipates or the two surfaces connect, the charge vanishes. Hence, the PZT can be viewed as a capacitor. When linked to a resistor, we obtain an RC circuit, enabling the conversion of the vibration into an electrical signal that can be measured. The design of the RC circuit is crucial for enhancing the quality of signal. The circuit utilized in our TapWristband is shown in Fig. 13. First, the resistance to the time constant of the sampling circuit $\tau = RC$ is considered. The $R1$ in the sampling circuit shouldn't be excessively large or small, as a high resistance impacts the voltage's rate of change, while a small resistance significantly reduces the sampling time, leading to distortion. In our practical system, $R1$ is set to $1M\Omega$. Second, the PZT may generate a negative voltage when it vibrates with the wrist. However, most ADCs are incapable of sampling negative voltages. To address this issue, we incorporate a DC bias circuit into the hardware system, as illustrated in Fig. 13.

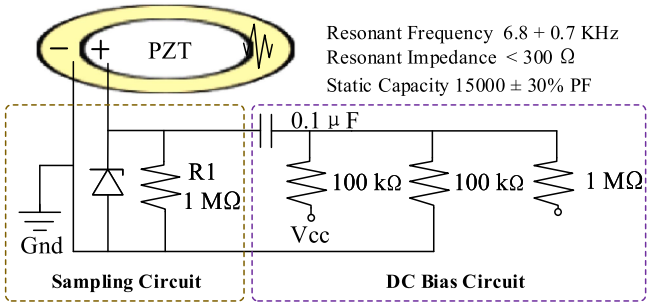

Fig. 13. Data Acquisition Circuit, where sampling circuit is used to collect the voltage signal from the vibration and the DC bias circuit is used to change the interval of the voltage signal.

C. Architectural Optimization

In our implementation, we employ a parallel structure for the acquisition and read processes. To ensure efficient operation on the chip, we extract only the signals within the window function for processing during the reading process. By maintaining two queues in the task to store the addresses of the extraction points on the streaming data, we can utilize shorter window functions to further enhance the operation speed without the need for a large amount of data.

Moreover, the temporal precision of the acquired signal influences the frequency accuracy and, consequently, the acquisition results. Given that the temporal stability of the acquired data is impacted by the reading, we stabilize the data sampling time. Our data is adjusted for time errors using an independent timer.

VII. Experiment and Evaluation

TapWristband is a potent and user-friendly system, making the assessment of its performance and reliability crucial. We initiate this process by detailing the experimental setup and proceeding with the system evaluation. As our work tackles a distinct problem compared to other related studies, we conduct a systematic evaluation of the system, focusing on aspects such as accuracy, keystroke rate, and robustness.

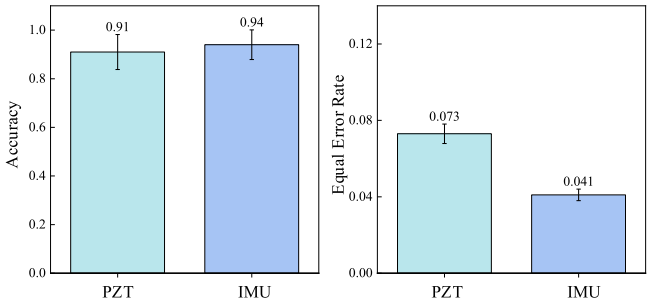

Fig. 14. Experimental results with different sensors.

A. Experimental Setup

We enroll a group of 17 volunteers, including 3 females, aged between 22 and 54, from our school. To validate the fundamental performance of TapWristband, all participants were asked to wear it on their right wrists. Our primary experiments were carried out in a typical indoor setting with ambient noise levels below 50dB. Participants were instructed to position their wrists over a table where a piece of paper, featuring a drawn keypad (as shown in Fig. 1), was placed. They were then asked to tap each key on the paper 10 times. We conducted three rounds of experiments in total, with a minimum interval of 1 h between each tapping session. After denoising and segmenting all signals, we obtained a total of over 8500 taps (30 taps per key per person) for training and testing. Among them, there are about 6000 sets of signals based on IMU sensors, and about 2500 sets of signals based on PZT sensors. To evaluate the robustness of our system and its performance under rapid tapping, some participants were asked to conduct supplementary experiments under different experimental conditions, which will be detailed in the following sections. All experiments commenced with the default settings unless otherwise specified.

B. Recognition Accuracy

In this section, we will discuss the basic performance of TapWristband, including its accuracy and error rate in recognizing different keystrokes. We also evaluate the impact of different sensors and each module of TapWristband on the experimental results.

1) Classification Accuracy: To evaluate the adaptability of the TapWristband's key recognition framework to various vibration sensors, we conducted experiments using two different sensors (IMU and PZT). As mentioned above, the data we collected contains about 5000 sets of data from 7 volunteers, half of which are from IMU sensors and half from PZT sensors. We divided them into two datasets according to the type of sensor, each with about 2500 sets of data. We used the model introduced in Section V to train and test these two datasets separately. The results of 10-fold cross-validation for the two datasets are shown in Fig. 14. The 'Accuracy' in Fig. 14 refers to the ratio of correctly identified samples to total samples, and 'equal error rate (EER)' refers to the ratio of incorrectly identified samples to total samples. The average recognition accuracy of the two groups of sensors is shown in Fig. 14. The results show that regardless of the type of sensor, the average recognition accuracy can exceed 90%. These results strongly prove that the recognition framework we proposed has broad applicability. We observe that the recognition accuracy is higher when using IMU, because the integration of IMU sensors is better,which makes the perception results more stable.

2) Effect of Sensor Position: In practical operations, we often encounter situations where the sensor position drifts and is not in the position we set. To investigate the tap recognition results when the sensors were in different positions, we conducted the following experiments. First, we placed the sensor at 5 positions as shown in Fig. 15(a), with a gap of $0.5cm$ between adjacent positions. Then we asked two participants (2 of the 7 volunteers) to press each key 10 times with different sensors, a total of 2 sets of 240 datasets as the test set. We used the model trained in the previous experiment to test these two test sets, and the final results of the average accuracy and KRR (key recognition rate) are shown in Fig. 15(b) and (c). The accuracy is the proportion of correctly identified keys to valid signals, and KRR represents the proportion of those recognized as a key to the total number of taps.

From the results, it can be seen that both PZT and IMU sensors need to ensure that the sensor is as far as possible in the mid-axis of the wrist, which will better ensure recognition. When the sensor is too close to the wrist, the sensor receives more interference from the raising and lowering of the wrist, resulting in a lower recognition rate and ultimately, accuracy. When the sensor is far away from the wrist, the signal attenuation is larger, which also leads to larger errors. In addition, a side-by-side comparison of the PZT and IMU sensors shows that the PZT is less effective due to the shape of the sensor, which does not ensure stable coupling. This is one of the reasons why the overall results for the PZT are slightly worse than the IMU in Fig. 15. Due to this advantage of IMUs, we followed up with experiments mainly using IMU sensors.

3) Confusion Matrix: To further evaluate the intrinsic performance of TapWristband, as mentioned in the previous question, we collected a total of about 6000 sets of data from 17 volunteers. We used 5000 of these as a training set and 1000 as a test set for validation. The confusion matrix in Fig. 16 shows the recognition accuracy of these participants. Our research results show that the recognition rate for each key exceeds 88%, with an average accuracy of over 93%. The confusion matrix further indicates that the discrimination ability of vertical keys is superior to that of horizontal keys, a phenomenon attributable to the significant structural difference between the fingertip and the finger belly. Given that the problem we proposed is different from the problems solved by existing research, we chose to use the aforementioned results as the baseline for the next evaluation. After training the model, it was used in subsequent experiments.

4) Effect of Tapping Position: A notable feature of our TapWristband is that the tapping is not affected by specific keyboard positions. To explore this feature, we conducted the following experiment: In the model trained in the previous section, we asked 2 subjects to tap 10 times at different positions on the keyboard with each finger according to the trained actions, and record the keys represented by their actions. It is important to

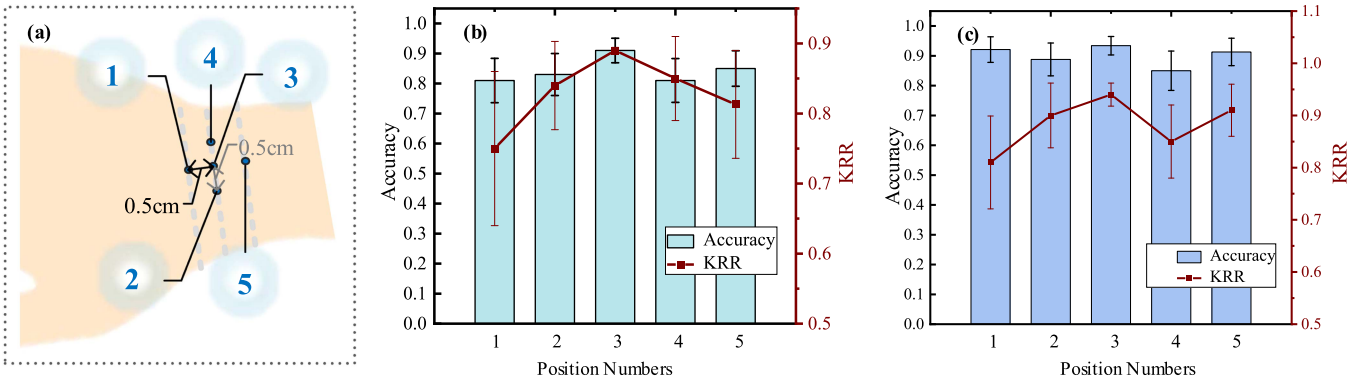

Fig. 15. Effect of different sensor wearing positions. (a) Schematic of different sensor wearing positions. (b) Effect of different sensor wearing positions on PZT sensors. (c) Effect of different sensor wearing positions on IMU sensors.

	Key0	Key1	Key2	Key3	Key4	Key5	Key6	Key7	Key8	Key9	Key10	Key11
Key0'	0.95	0	0	0	0	0	0	0	0	0	0	0
Key1'	0	0.96	0.02	0	0	0.01	0	0	0	0	0.01	0
Key2'	0	0	0.9	0	0	0	0	0	0	0	0	0
Key3'	0	0	0.03	0.94	0	0	0.05	0	0	0.01	0.03	0
Key4'	0	0	0	0	0.91	0.01	0	0.01	0	0	0	0
Key5'	0	0	0	0	0.02	0.93	0.03	0	0	0	0	0
Key6'	0	0	0	0	0.02	0.01	0.88	0	0	0	0	0
Key7'	0.01	0	0	0	0.02	0	0	0.92	0	0	0	0
Key8'	0	0	0.01	0	0.01	0.03	0	0.02	0.96	0.01	0.01	0.01
Key9'	0	0	0.01	0.01	0.01	0	0.01	0.02	0	0.93	0	0
Key10'	0	0	0	0	0	0	0	0	0	0	0.88	0
Key11'	0	0	0	0	0.01	0	0	0	0	0.01	0.04	0.96

Fig. 16. Confusion matrix of 12 keys, the x-axis represents the actual keystrokes and the y-axis represents the output of TapWristband.

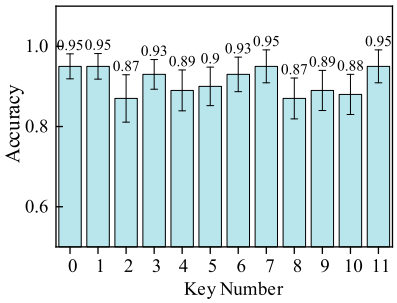

Fig. 17. Relatively stable gesture results for clicks at different key positions.

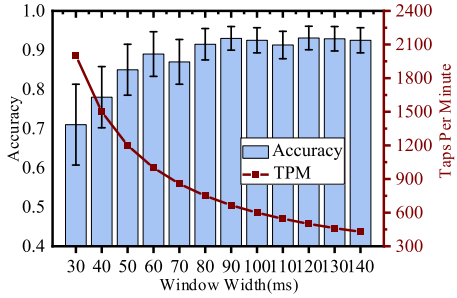

Fig. 18. Effect of energy window width L on accuracy and TPM.

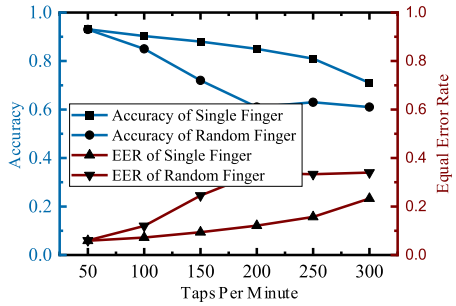

Fig. 19. Effect of rapid tapping on accuracy and EER.

note that the subject's seat remains fixed while tapping, the keyboard is panned across the desktop, and the relative position required for each key is maintained as the key is tapped (e.g., the key 2 is tapped using the belly of the middle finger). We obtained 240 sets of data as a test set, and the average accuracy results after multiple taps are shown in Fig. 17. The results in the figure verify that TapWristband has the ability to input flexibly without restrictions.

5) Bandwidth of Filter: Since a narrowband filter is used for signal shaping in the decay fitting process of Algorithm 1. Therefore, we investigated the effect of filter bandwidth on the Accuracy and signal recognition rate. We define 'effectiveness' as the ratio of the number of signals recognized as tap to the total number of signals. When the bandwidth of the filter is too narrow, it ensures the relative purity of the signal but also removes the effective information. This results in some noise being forced to be filtered out as valid information, which ultimately leads to low accuracy.

Of course, a filter with too wide a bandwidth will result in events that cannot be effectively filtered and thus valid events that are perturbed by noise cannot be accurately extracted, as shown in Fig. 20.

6) Ablation Experiments: Subsequently, we carried out ablation experiments to examine the importance of various designs in both event recognition and keystroke extraction. As depicted in Fig. 21 B, the average accuracy rate of recognition is presented

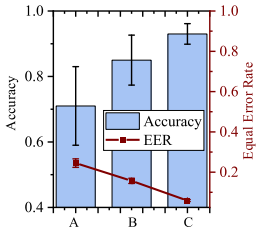

Fig. 20. Ablation Experiment, where A represents the result of no decay fitting and precise extraction. B represents the result of no precise extraction. C represents the baseline.

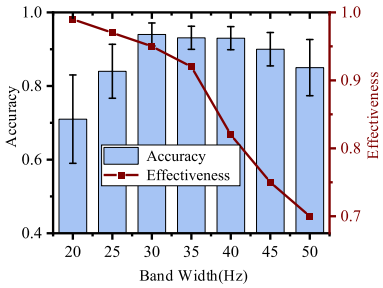

Fig. 21. Effect of Butterworth filter bandwidth on Accuracy and Effectiveness.

after the removal of the precise extraction module based on mutual correlation. The lack of precise segmentation of the signal results in the mixing of forced vibrations, causing significant disruption to the recognition process. Fig. 21 A illustrates the scenario when the fading monitoring is removed on top of B. This leads to incomplete filtering of events and the introduction of noise into the dataset, resulting in substantial interference.

C. Rapid Tapping

Input efficiency is a crucial metric in the assessment of an HCI system. In this subsection, we delve into the primary elements influencing the recognition speed of TapWristband. Additionally, we evaluate the impact of rapid tapping on the accuracy of TapWristband.

1) Segment Evaluation: We conducted a study to evaluate the impact of the segmentation window length L on the system's accuracy and maximum input efficiency. We used the data and model from Section VII-B3) for this experiment, incrementing L from 30 to 140 and calculating the system's average accuracy of the system at each increment. The research results are shown in Fig. 18. These results indicate that a window that is too short will result in the loss of some signal information. In addition, a window that is too long will introduce additional noise to the system, and the number of taps per minute (TPM) is inversely proportional to the window length. Therefore, unless otherwise specified, we set $L = 80$ for all subsequent research based on the above conclusions.

2) Tap Speed Influence: Rapid tapping can lead to inevitable muscle tension, which affects the system's recognition accuracy. To verify the rapid capability of TapWristband and its proficiency in handling high-efficiency, large-capacity tapping tasks, we conducted experiments on two volunteers. A metronome was used in the experiment to synchronize the tapping rhythm of the volunteers' fingers with the rhythm of the metronome, striking about 1 time per beat points with a total of about 100 taps at a fixed rhythm. The experimental results are shown in Fig. 19 of the novice manuscript, with the system's accuracy decreasing as the number of finger taps per minute increases. Fig. 18 shows that although the volunteers' tapping speed is significantly lower than the theoretical maximum speed, continuous forced movement still affects relaxing vibration.

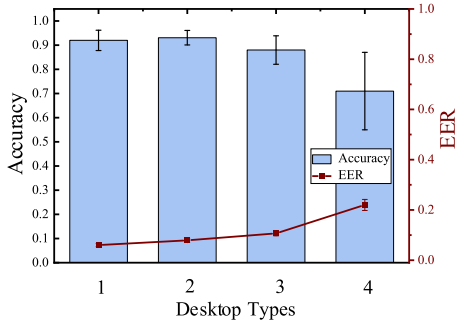

Fig. 22. Effect of Firmness of the table on accuracy and EER.

Among the volunteers who participated in the experiment, their typing speed is approximately 120-200 characters per minute. In this survey, they reported that when using our technology for continuous typing, the highest tapping speed remains at 120-140 times per minute. As shown in Fig. 19 of the novice manuscript, even if tapping 150 times per minute, the system's evaluated accuracy is close to 90%. It can be seen that although our technology may limit some people who type very fast on traditional keyboards, it can meet the input needs of medium speed.

D. Robustness

In this subsection, we investigate the robustness of TapWristband under some real disturbances, including ambient noise, vibration, and so on.

1) Firmness of the Table: We tested the impact of the hardness of the table on the final recognition accuracy. Using the previous model, we asked two volunteers to tap each key ten times on each different surface of the table as a test set. The results are shown in Fig. 22. Here, 1 represents a stone desktop, 2 represents a wooden desktop, 3 represents a desktop with a soft pad, and 4 represents a false point press. It can be seen that our technology has certain requirements for the hardness of the desktop, and the performance is not ideal when the tapping is on a void point or on a soft desktop.

2) Environmental Noise Influence: We wanted to explore whether TapWristband would still work for the noise environment. To do this, we first investigated the accuracy of TapWristband in the presence of varying levels of ambient noise. In this experiment, we recruited three volunteers and let them tap each key 10 times under different ambient noises. The classification model was used as mentioned above, which was trained in advance. As can be seen from Fig. 23, the accuracy of TapWristband decreases not very significantly as the external

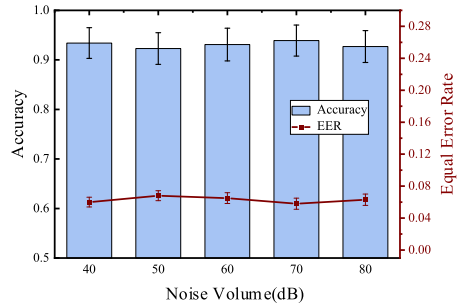

Fig. 23. Effect of ambient sound on accuracy and EER.

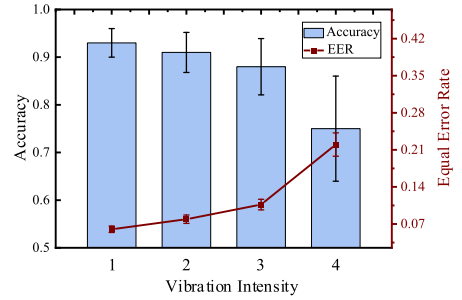

Fig. 24. Effect of desktop vibration on accuracy and EER.

sound continues to amplify, indicating that TapWristband has strong robustness in terms of sound.

Since TapWristband mainly detects the vibration of the wrist to recognize the keystrokes, we further investigated the effect of desktop vibration on the accuracy of TapWristband. We set up some scenarios that can be seen in daily life in terms of vibration, and the other experimental settings are consistent with the above. We classify the vibration generated by the desktop in daily life into four levels, the first level is the vibration generated by the friction of the body and clothes in daily life; the second level is the vibration generated by some desktop daily products, such as laptops, fans, and mainframes; the third level is a kind of sudden vibration, such as vibration generated by cell phone calls, alarm clock vibration, and so on; the fourth level is the vibration generated by daily vibration that can cause discomfort to the human being such as someone knocking or kicking the table.

The results are shown in Fig. 24, it can be seen that at the first and second vibration levels, the recognition accuracy is maintained at more than 90% correct although there is a decrease. As the vibration is strengthened, the accuracy of TapWristband decreases rapidly, and the EER rises rapidly. However, this strong vibration is short-lived, so it can be said that TapWristband is robust to vibration in everyday situations.

3) Long-Term Performance: As a system dedicated to providing a simple and convenient user experience, TapWristband aims to evaluate the long-term performance of users in the absence of keyboard prompts. To this end, we have integrated the data of a long-term user into the data model of several existing participants.

This long-term user did not use the paper required for the experiment, but randomly clicked each key about 10 times. After a few hours, we asked the user to tap again, and conducted three sets of experiments every day for the next five days. After denoising and segmentation, we selected more than 1400 keys for training and testing. We used a 10-fold stratified cross-validation method for evaluation.

TABLE II
PERFORMANCE OF LONG-TERM LEARNING

Usage Period(Day)	Accuracy(%)	EER(%)	Tapping Speed(TPM)
1	88.2	8.9	55.9
2	90.9	7.1	68.5
5	93.2	5.3	101.2
10	93.3	5.0	107.1
30	93.1	5.1	105.3
90	92.3	5.7	106.1
180	90.5	6.9	109.3

TABLE III
SURVEY OF USERS

Items	Comfort	Efficiency	Traning&Learn
Gesture	Yes	Low	Positive
Typing	Medium	Medium	Positive
TapWristband	Yes	Positive	Medium

The tap recognition results of this long-term user are shown in Table II. In less than a week, the participant was already able to use TapWristband quite well. To further evaluate its long-term performance, we deliberately extended the usage time, and then conducted experiments after one month, three months, and six months, the results of which are shown in Table II. These results clearly show that TapWristband has good durability in long-term use. Furthermore, we shuffled the six-month data of the subjects into the models of the aforementioned seven individuals, conducted a 10-fold cross-validation, and ultimately achieved an average accuracy rate of over 92%. This gives us confidence to continue to optimize and improve TapWristband to meet the long-term usage needs of users.

VIII. DISCUSSION

In this section, we discussed the design objectives of TapWristband, user research, energy consumption, outlook on integrated applications, and future work.

A. User Study

One of the main objectives of TapWristband is to achieve universal applicability and user-friendliness. To realize this goal, we proactively collect user feedback after they have used it. This feedback mechanism allows us to better understand the needs and expectations of the users, thereby improving our technology.

In our research, we conducted a comparative study involving 17 participants, where we evaluated the features of TapWristband against the touchscreen gesture control and typing functionalities of a smartwatch, as shown in Table III. The majority of participants expressed a preference for the comfort provided by the gesture controls and the use of TapWristband, while the typing feature received mixed feedback. When assessing input efficiency, TapWristband emerged as the most efficient, followed

by typing, with gesture control being the least efficient. Participants generally expressed a positive attitude towards learning to use these three features. However, a subset of participants noted that training for TapWristband was more time-intensive.

Regarding the issue reported by some users about the training portion being somewhat slow, this is very important feedback. We will take it into consideration in future versions to enhance the efficiency and effectiveness of the training.

B. Power Optimization

For a system powered by batteries, energy consumption is an important issue. Currently, the energy consumption of the whole system consists of several main components classified according to the system architecture: power consumed by the sensors, power consumed by the analogue-to-digital converter, and power consumed by the processor. The energy consumption of sensing and acquisition is highly related to the data return rate, and the frequency of acquisition in our system is not high. The processor's energy consumption mainly depends on the number of operations. In this case, optimizing the inference method could significantly improve energy efficiency.

Incorporating a hibernation mechanism is a simple and effective way to save energy, in addition to using devices and methods that consume less energy. This involves turning the system on while in use and letting it hibernate when not in use. We can work with the chip's hibernation feature to wake up the system by using vibration as a trigger. It is worth noting that the PZT sensor, as a passive device, provides a great advantage by allowing the system to remain in a low-power state as much as possible. In our future work, we will explore battery optimization techniques from the points mentioned above.

C. Integration and Application

The TapWristband framework, defined by two main components - the sensor above the wrist and the hard object below the hand, can be easily integrated into many different systems. We divide the integration of TapWristband with existing devices into two categories, signalling and command transmission. For wrist devices such as smartwatches, TapWristband can be easily integrated into them by simply reusing the framework described in the content of this paper to develop the app and configure the responsive sensors in the system. The framework in this paper does not limit the output of signals, and thus the system can be used to do whatever the user wants to do with it, whether it is for keying in information, controlling the system, or shortcutting commands. For other smart devices, as we are currently doing, a bracelet can be configured to deploy the aforementioned application and sensors, enabling it to be used as a keyboard. This keyboard can be linked to the smart device in any way needed, such as Bluetooth, USB, and other protocols.

D. Future Work

In our future work, we plan to continue to explore our system, including investigating a variety of sensors and the effect of the sensor sampling rate on the system. We also plan to explore the possibility of using multiple vibration sensors, e.g., IMU and PZT sensors at the same time. We believe that by integrating data from IMU and PZT sensors, we can further improve the performance of our system. This multi-sensor approach can provide more useful information, helping us to identify and interpret user taps more accurately.

In addition, we plan to collect more user data to further optimize and expand our model. By using techniques including but not limited to transfer learning and incremental learning, we aim to improve the efficiency of our system and the user experience, enabling new users to start using the system immediately without waiting for model training to complete.

In summary, the future work mentioned above can lead to improvements making our TapWristband system more in line with user needs and playing an important role in the field of human-computer interaction.

IX. Conclusion

In this study, we propose TapWristband, a system that uses vibration sensors on wearable devices to detect and classify finger tapping, thereby enabling text input on any surface and effectively realizing a virtual keypad. We perform real-world experiments to collect measurements for modeling the effects of the finger-tapping motion on wearable wristband sensors. We develop pre-processing, tap detection, vibration segmentation, feature extraction, and classification algorithms to recognize the tapping patterns of five fingers across twelve key locations of a keypad system. Finally, we performed extensive experiments with thirteen participants to evaluate our system. Our experimental results demonstrate that the system is accurate (with an accuracy rate exceeding 93%), dependable, and efficient in practical applications. It can be seamlessly incorporated into various smart wristbands at low cost. Our initial findings suggest that TapWristband has significant potential in using wearable IoT sensors for HCI applications.

References

[1] S. Park and S. Jayaraman, "Chapter 1 - wearables: Fundamentals, advancements, and a roadmap for the future," in *Wearable Sensors*, 2nd ed., E. Sazonov ed., New York, NY, USA: Academic Press, pp. 3–27, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128192467000012

[2] E. Lamarre and B. May, *Ten Trends Shaping the Internet of Things Business Landscape*, New York, NY, USA: McKinsey Digital, 2019.

[3] E. Jones, J. Alexander, A. Andreou, P. Irani, and S. Subramanian, "Gestext: Accelerometer-based gestural text-entry systems," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2010, pp. 2173–2182.

[4] S. C. A., Y. L. B., Z. T. C., W. C. D., and X. C. E., "A model for integrating heterogeneous sensory data in IoT systems," *Comput. Netw.*, vol. 150, pp. 1–14, 2019.

[5] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "FingerIO: Using active sonar for fine-grained finger tracking," in *Proc. Chi Conf.*, 2016, pp. 1515–1525.

[6] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, 2016.

[7] S. Cheng, J. Yan, J. Li, and J. Liu, "Typingwristband: A human slight motion sensing system based on vibration detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2021, pp. 8313–8317.

[8] W. Chen et al., "ViType: A cost efficient on-body typing system through vibration," in *Proc. 15th Annu. IEEE Int. Conf. Sens., Commun., Netw.*, 2018, pp. 1–9.

[9] R. Xiao, G. Lew, J. Marsanico, D. Hariharan, S. Hudson, and C. Harrison, "Toffee: Enabling ad hoc, around-device interaction with acoustic time-of-arrival correlation," in *Proc. 16th Int. Conf. Hum.-Comput. Interact. Mobile Devices Serv.*, New York, NY, USA, 2014, pp. 67–76. [Online]. Available: https://doi.org/10.1145/2628363.2628383
[10] E. Miluzzo, A. Varshavsky, S. Balakrishnan, and R. R. Choudhury, "Tapprints: Your finger taps have fingerprints," in *Proc. Int. Conf. Mobile Syst., Appl., Serv.*, 2012, pp. 323–336.
[11] S. Nirjon, J. Gummeson, D. Gelb, and K.-H. Kim, "TypingRing: A wearable ring platform for text input," in *Proc. 13th Annu. Int. Conf. Mobile Syst., Appl., Serv.*, New York, NY, USA, 2015, pp. 227–239. [Online]. Available: https://doi.org/10.1145/2742647.2742665
[12] H. Dekker, "Classical and quantum mechanics of the damped harmonic oscillator," *Phys. Rep.*, vol. 80, no. 1, pp. 1–110, 1981.
[13] W. Chen, L. Chen, Y. Huang, X. Zhang, and K. Wu, "Taprint: Secure text input for commodity smart wristbands," in *Proc. 25th Annu. Int. Conf.*, 2019, pp. 1–16.
[14] J. Liu, Y. Chen, M. Gruteser, and Y. Wang, "VibSense: Sensing touches on ubiquitous surfaces through vibration," in *Proc. 14th Annu. IEEE Int. Conf. Sens., Commun., Netw.*, 2017, pp. 1–9.
[15] A. S. Rathore et al., "Sonicprint: A generally adoptable and secure fingerprint biometrics in smart devices," in *Proc. 18th Annu. Int. Conf. Mobile Syst., Appl., Serv.*, 2020, pp. 121–134.
[16] C. Harrison, D. S. Tan, and D. Morris, "Skinput: Appropriating the body as an input surface," in *Proc. 28th Int. Conf. Hum. Factors Comput. Syst.*, Atlanta, Georgia, USA, 2010, pp. 453–462.
[17] Rabiner and R. L., "A tutorial on hidden markov models and selected applications in speech recognition," in *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
[18] G. Hinton et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
[19] C. Liang, C. Yu, Y. Qin, Y. Wang, and Y. Shi, "DualRing: Enabling subtle and expressive hand interaction with dual IMU rings," in *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 5, no. 3, pp. 115:1–115:27, Sep. 2021.
[20] S. McKinley and M. Levine, "Cubic spline interpolation," *College Redwoods*, vol. 45, no. 1, pp. 1049–1060, 1998.
[21] J. Huang et al., "A universal and arbitrary tactile interactive system based on self-powered optical communication," *Nano Energy*, vol. 69, 2020, Art. no. 104419. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211285519311346
[22] M. Meier, P. Streli, A. Fender, and C. Holz, "TapID: Rapid touch interaction in virtual reality using wearable sensing," in *Proc. IEEE Virtual Reality 3D User Interfaces*, 2021, pp. 519–528.
[23] W. Chen et al., "Robust finger interactions with cots smartwatches via unsupervised siamese adaptation," in *Proc. 36th Annu. ACM Symp. User Interface Softw. Technol.*, New York, NY, USA, 2023, pp. 25:1–25:14. [Online]. Available: https://doi.org/10.1145/3586183.3606794
[24] C. Zhang et al., "FingerPing: Recognizing fine-grained hand poses using active acoustic on-body sensing," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Montreal QC Canada, 2018, pp. 1–10.
[25] H. Zhou, T. Lu, Y. Liu, S. Zhang, and M. Gowda, "Learning on the rings: Self-supervised 3d finger motion tracking using wearable sensors," in *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 6, no. 2, pp. 90:1–90:31, Jul. 2022. [Online]. Available: https://doi.org/10.1145/3534587
[26] R. Hajika et al., "Radarhand: A wrist-worn radar for on-skin touch-based proprioceptive gestures," *ACM Trans. Comput.-Hum. Interact.*, vol. 31, pp. 1–36, Oct. 2023. [Online]. Available: https://doi.org/10.1145/3617365
[27] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proc. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 82–94.
[28] F. Khan, S. K. Leem, and S. H. Cho, "In-air continuous writing using UWB impulse radar sensors," *IEEE Access*, vol. 8, pp. 99302–99311, 2020.
[29] Y. Huang, C. Li, F. Chen, Q. Zhang, and K. Wu, "MM-tap: Adaptive and scalable tap localization on ubiquitous surfaces with MM-level accuracy," *IEEE Internet Things J.*, vol. 10, no. 23, pp. 20789–20802, Dec. 2023.
[30] X. Xu et al., "Enabling hand gesture customization on wrist-worn devices," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2022, pp. 1–19.
[31] J. Shang and J. Wu, "Fine-grained vital signs estimation using commercial Wi-Fi devices," in *Proc. 8h Wireless Students, By Students, Students Workshop*, 2016, pp. 30–32.
[32] S. Lee, W. Choi, and D. H. Lee, "Usable user authentication on a smartwatch using vibration," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2021, pp. 304–319.
[33] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range FMCW monopulse radar for hand-gesture sensing," in *Proc. IEEE Nat. Radar Conf. - Proc.*, 2015, pp. 1491–1496.
[34] D. Li, J. Liu, S. I. Lee, and J. Xiong, "LASense: Pushing the limits of fine-grained activity sensing using acoustic signals," in *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 6, no. 1, pp. 1–27, Mar. 2022. [Online]. Available: https://doi.org/10.1145/3517253
[35] R.-D. Vatavu, "iFAD gestures: Understanding users' gesture input performance with index-finger augmentation devices," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2023, pp. 1–17. [Online]. Available: https://doi.org/10.1145/3544548.3580928
[36] A. Byanjankar, Y. Liu, Y. Shu, I. Shin, M. Choi, and H. Kim, "S-UbiTap: Leveraging acoustic dispersion for ubiquitous and scalable touch interface on solid surfaces," *IEEE Trans. Mobile Comput.*, vol. 22, no. 11, pp. 6800–6816, Nov. 2023.
[37] S. Pan, C. G. Ramirez, M. Mirshekari, J. Fagert, and P. Zhang, "Surfacevibe: Vibration-based tap & swipe tracking on ubiquitous surfaces," in *Proc. ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, 2017, pp. 197–208.
[38] L. Cai and H. Chen, "TouchLogger: Inferring keystrokes on touch screen from smartphone motion," in *Proc. 6th USENIX Conf. Hot Topics Secur.*, ser. HotSec'11, USA, 2011, p. 9.
[39] E. Markvicka, G. Wang, Y.-C. Lee, G. Laput, C. Majidi, and L. Yao, "ElectroDermis: Fully untethered, stretchable, and highly-customizable electronic bandages," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2019, pp. 1–10.
[40] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.
[41] Z. Xu et al., "TipText: Eyes-free text entry on a fingertip keyboard," in *Proc. 32nd Annu. ACM Symp. User Interface Softw. Technol.*, New York, NY, USA, 2019, pp. 883–899.
[42] Y. Iravantchi, Y. Zhang, E. Bernitsas, M. Goel, and C. Harrison, "Interferi: Gesture sensing using on-body acoustic interferometry," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, 2019, pp. 1–13. [Online]. Available: https://doi.org/10.1145/3290605.3300506
[43] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
[44] Antoniou, *Digital Filters, Analysis, Design, and Applications*. New York, NY, USA: McGrawHill, 1993.
[45] M. B. Kursa and W. R. Rudnicki, "Feature selection with the boruta package," *J. Statist. Softw.*, vol. 36, pp. 1–13, 2010.

Jialiang Yan received the BS and master's degrees in electronic science and technology from the Harbin Institute of Technology. He is currently working toward the PhD degree with the School of Computer Science and Technology, Harbin Institute of Technology. His research interests include wireless sensing, human-computer interaction and wireless communication.

Siyao Cheng received the BS, master's, and PhD degrees in computer science from he Harbin Institute of Technology. She is a professor with the School of Computer Science and Technology, Harbin Institute of Technology. Her research interests include big sensory data management, wireless sensor networks and edge computing.

Yang Zhao (Senior Member, IEEE) received the BS degree in electrical engineering from Shandong University in 2003, the MS degree in electrical engineering from the Beijing University of Aeronautics and Astronautics in 2006, and the PhD degree in electrical and computer engineering from the University of Utah in 2012. He was a lead research engineer with GE Global Research between 2013 and 2021. Since 2021, he has been with the Harbin Institute of Technology, Shenzhen, where he is a research professor with the International Research Institute for Artificial Intelligence. His research interests include wireless sensing, edge computing and cyber physical systems.

Jie Liu (Fellow, IEEE) is a chair professor with the Harbin Institute of Technology (Shenzhen), China and the dean of its AI Research Institute. He is an ACM distinguished scientist. Before joining HIT, he spent 18 years with Xerox PARC, Microsoft Research and Microsoft product teams. As a principal research manager with MSR, he led the Sensing and Energy Research Group (SERG). In MSR-NExT and product groups, he incubated smart retail solutions, which became part of Microsoft Business AI offering. His research interests root in understanding and managing the physical properties of computing. He also has chaired a number of top-tier conferences in sensing and pervasive computing, and was associate editors for several top-tier journals.