ViType: A Cost Efficient On-body Typing System through Vibration

Wenqiang Chen*, Maoning Guan†, Yandao Huang*, Lu Wang*, Rukhsana Ruby*, Wen Hu‡ and Kaishun Wu*

*College of Computer Science and Software Engineering, Shenzhen University
†College of Information Engineering, Shenzhen University
‡University of New South Wales, Australia

***Abstract*—Nowadays, smart wristbands have become one of the most prevailing wearable devices as they are small and portable. However, due to the limited size of the touch screens, smart wristbands typically have poor interactive experience. There are a few works appropriating the human body as a surface to extend the input. Yet by using multiple sensors at high sampling rates, they are not portable and are energy-consuming in practice. To break this stalemate, we proposed a portable, cost efficient text-entry system, termed ViType, which firstly leverages a single small form factor sensor to achieve a practical user input with much lower sampling rates. To enhance the input accuracy with less vibration information introduced by lower sampling rate, ViType designs a set of novel mechanisms, including an artificial neural network to process the vibration signals, and a runtime calibration and adaptation scheme to recover the error due to temporal instability. Extensive experiments have been conducted on 30 human subjects. The results demonstrate that ViType is robust to fight against various confounding factors. The average recognition accuracy is 94.8% with an initial training sample size of 20 for each key, which is 1.52 times higher than the state-of-the-art on-body typing system. Furthermore, when turning on the runtime calibration and adaptation system to update and enlarge the training sample size, the accuracy can reach around 98% on average during one month.**

I. Introduction

In the past few years, we have seen the take-off of wearable wristbands such as Fitbit and Apple iWatch for fitness applications. People begin to use more applications such as electronic payment and short message service(SMS) on smart wristbands instead of mobile phones. The size of smart wristbands have become smaller and lighter to provide better user experience. As a result, the touch screens on the wristbands also become smaller, while human fingers do not shrink accordingly, which are difficult to support text input.

Currently, to overcome the limitations of a small screen, speech recognition is one of the methods but is sensitive to noise levels in the surrounding environments. Moreover, it is insecure for sensitive information (e.g., password input) since speech input is easy to be eavesdropped. For the same reason, it is also intrusive to the people surrounding the user. Recent works by FingerIO [1] and LLAP [2] achieves millimeter-scale localization accuracy for fingertip tracking, which enables users to write letters on ubiquitous surfaces instead of tiny touch screens. However, writing letters is significantly slower than typing them, which results in poor user experience.

Fig. 1. A sample example of ViType.

In this paper, we present a novel system termed ViType, as shown in in Fig. 1, which enables a user to type on the back of one's hands (opisthenar) instead of a tiny touch screen of a smart wristband. We leverage a vibration sensor on a wrist to collect the vibration signal by tapping in different locations on the opisthenar. As the signal carries diverse energy at different frequencies and over different distances, we reap its benefit as the unique input feature. ViType inherits the merits of vibration, such as resistance to acoustic noise and environmental dynamics. Moreover, the size of the opisthenar is larger than tiny touch screens, which enables the user to type more quickly and conveniently.

Motivated by this, we have designed a keystroke recognition system that leverages location-based vibration information derived from a small piezoelectric ceramic vibration sensor. The sensor can be easily embedded to a smart wristband. Although keystroke recognition via body vibration has been studied in Skinput [3], it takes 10 sensors of an armband with a very high sampling rate (e.g., 55 kHz) to collect the signal. On the contrary, as depicted in Fig. 2, ViType only uses a single small form factor sensor which makes it easier and more cost effective. Furthermore, it samples at lower sampling rates (e.g., 600 Hz) to make it more efficient for running on resource limited smart wristbands.

978-1-5386-4281-8/18/$31.00 ©2018 IEEE

It is nontrivial to embrace the above vision, as sampling at a lower rate produces significantly less vibration information. Hence, we need to investigate novel methods for keystroke recognition via body vibrations. Furthermore, in our vision, ViType should not only attain high accuracy but also be robust to many practical issues. For example, although ViType is a location-based training system, users' wristbands may have a little displacement while typing over the time. Second, users may type with different force or different finger/hand posture. Third, ViType is expected to be functional when users are walking which may cause vibration noise to the system. Fourth, it should be convenient for users to train the system at the time of first usage and then use it successively.

To cope with these challenges, we studied a set of novel keystroke detection/ classification mechanisms on different vibration patterns produced by keystrokes on users' opisthenar. We find that keystroke recognition scheme via location-based vibration depends on the vibration amplitude and frequency, which can be characterized by the waveforms and power spectral density (PSD). A more important observation is that the waveforms and PSD of different keystroke locations reveal highly distinguishable profiles and can be conveniently used as a location signature. We removed the noise signal caused by human mobility from the original signal via a filter and then used an online dual-threshold start point detection algorithm to detect keystroke signals. In addition, we find that Artificial neural network (ANN) with a min-max normalization concept is a suitable technique for vibration classification. A regularization is further employed particularly due to the limited size of training samples. Last but not the least, we design a runtime calibration and adaptation system and provide a special scheme to update and enlarge the training set to enhance the robustness in practical situations such as variations of finger posture or displacement of wristbands and tap position.

We implement ViType as a prototype on a Raspberry Pi with a small form factor piezoelectric ceramic in real time system [21]. A demonstration video is attached in the link[1]. Our baseline evaluation shows that classification accuracy is 94.8% on average for 30 human subjects with an initial training sample size of 20 for each key, which is 1.52 times higher than the state-of-the-art system of Skinput [3]. We have also conducted a series of studies in realistic settings such as wristband or tapping displacement, variations of tapping force, and found that the performance of ViType degrades significantly in non-ideal circumstances. Thus, we design a runtime calibration and adaptation scheme to address these challenges and our results show that the proposed scheme can mitigate the degradation.

Our contributions in this work lie in the following aspects.

- ViType is the first attempt in the literature to recognize the keystrokes typing on a user's opisthenar via a single small size vibration sensor. It samples at an order of magnitude lower rates to achieve a more efficient text-input method on resource limited smart wristbands.

[1]https://youtu.be/taLZLFyPB4M

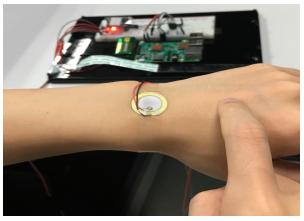

Fig. 2. A sample prototype of ViType.

- We president the entire design of ViType, which utilizes ANN to recognize the keystroke vibrations, and harnesses a runtime calibration and adaptation scheme to achieve a desirable recognition accuracy.
- We comprehensively evaluate the performance of ViType under different scenarios. The recognition accuracy is 94.8%, which is 1.52 times higher than the state-of-the-art on-body typing system.

The remainder of this paper is structured as follows. In Section II, we first provide the background information and the related work in the context of this work. Then, Section III presents the overview of ViType showing the design goals and challenges. Section IV describes the three main modules of ViType. Section V explains the detailed implementation technique, followed by a comprehensive experimental evaluation of our system. Finally, Section VI draws a conclusion on this paper.

II. Background and Related Work

Body vibration: Keystroke recognition via body vibration has been studied in Skinput [3], in which a signal is collected from 10 sensors of an armband with a very high sampling rate (e.g., 55 kHz). When a finger taps on the opisthenar, two separate forms of vibration are produced, which are transverse waves and longitudinal waves. Transverse waves translate along the hand surface while longitudinal waves move into and out of the bone through soft tissues. Moreover, during the propagation of vibration from a tapped location to sensor, the signals suffer attenuation and the model can be stated as follows [4].

$$A(d) = A_0 e^{-\alpha \times d}, \tag{1}$$

where A_0 is the initial amplitude, d is the propagation distance and α is the attenuation coefficient. The relation in (1) further reveals that the amplitude of vibration signal is dominated by the propagation distance and attenuation coefficient. Thus, vibrations resultant from tapping in different locations of opisthenar are distinct as they carry the diminishing energy at different frequencies over different distances. Additionally, higher frequencies propagate more readily through bone than through soft tissue, and bone conduction carries energy over larger distances compared to soft tissue [3]. As for attenuation coefficient, it is associated with the medium which vibration signals propagate through, which means it may vary in accordance with different physiological states of the human body

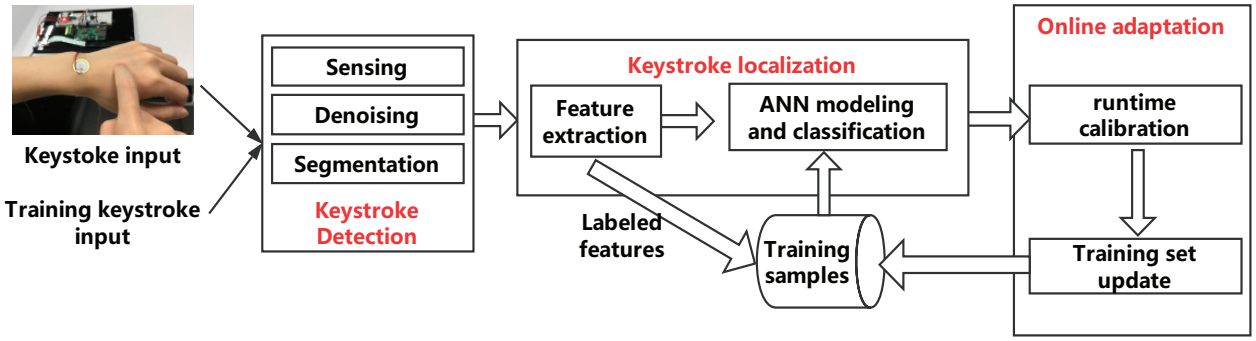

Fig. 3. Architecture of ViType

over time (Section V-C will discuss the temporal stability of ViType).

Text input for wearable devices: Speech recognition [5] is popular for text input. Furthermore, there are also some acoustic interface-based systems for text input [6][22]. However, it is sensitive to noise levels in the environments and insecure for sensitive information such as password input due to its convenience to be easily eavesdropped. For the same reason, it is also intrusive to the people around the user. Recent works by FingerIO [1] and LLAP [2] achieve millimeter-scale localization accuracy for fingertip tracking, which enables users to write letters on ubiquitous surfaces instead of tiny touch screens. However, writing letters is significantly slower compared to typing, which results in poor user experience.

Some projects require a user to carry extra devices such as a ring, a pen or even a shoulder-mounted camera for text input [7-11]. However, they are uncomfortable and cumbersome. It is also unacceptable for users to implant sensors under the skin [12] for text input. Camera technique suffers great controversy in terms of privacy issues and requires line-of-sight sensing [7]. Although infrared technique has achieved some simple human-computer interaction designs [13,14], it is still unsuitable for text input because of power consumption and accuracy limitation.[23-28]

Vibration sensors: Serendipity [15] leveraged accelerometer and gyroscope to recognize five finger gestures, such as pinching, tapping, rubbing, squeezing and waving. These accelerometers and gyroscopes are only used for recognizing finger gesture, but not adequate for keystrokes as these sensors are not sensitive to the vibration of finger taps. There are few works which used piezoelectric vibration sensors to classify keystrokes [16]. Recently, SurfaceVibe [17] used geophone to localize taps and swipes for supporting text input on surfaces such as wood tables. However, these methods used TDOA-based localization concept which require multiple sensors, and hence they are not fit for small size wearable wristbands. VibSense [18] also realizes the goal for typing on external surfaces such as wood table via vibration, but it generates vibration on a table by a vibrator while we sense vibration on body directly generated by finger which is more unstable and complex to be processed.

III. Overview of ViType

A. Design goals and challenges:

We design ViType to meet the following goals and overcome the following challenges which are basically required to use this system in practice.

1) User friendly: It will result in terrible experience if users have to reset the input system each time they type. Such time overhead is not negligible and annoying if the usage duration is short. Therefore, ViType needs to make sure its temporal stability that each user has to launch the setup procedure only once.

2) Availability: The localization mechanism of ViType should also be designed to fit in different operating conditions. For instance, users may apply different tapping force and change their hand posture over time and even need to type while in the walking phase. Besides, ViType should have a strong ability of resisting acoustic noise caused from surrounding environments.

3) Fine-grained: ViType utilizes the relatively wide area of the opisthenar as the interactive interface for users to type in. However, the space between two keys is only around 2 cm. In order to achieve a centimeter-scale keystroke localization, we have to realize a localization mechanism that can recognize keystrokes with high accuracy. In Skinput [3], the accurate localization on forearm is realized using 10 sensors of an armband at a very high sampling rate such as 55 kHz. However, ViType needs to achieve accurate localization on opisthenar using only one sensor at a low sampling rate such as 600 Hz.

4) Deviation-tolerant: Marking keyboard layout on the opisthenar with a pen is inconvenient and the marks on the hand tend to be erased, which will result in poor user experience. One of the schemes to perform a virtual keyboard on a particular interactive surface is projection [13], which can

visualize the layout of keyboard. However, it is impractical to project a virtual keyboard on the opisthenar as smart wristband is energy limited and extra employment of hardware is needed. Hence, ViType has to attain its localization with high accuracy in the circumstances that users have no visual keyboard layout and assistant marker to keep tapping on the same key position over time. In addition, similar to the deviation of keystrokes, the shift of a smart wristband over time also needs to be taken into consideration.

B. System architecture:

The architecture of ViType consists of three major components in order to build a robust and self-contained keystroke localization system for smart wristband. The functionalities of these components are described in the following.

1) Keystroke detection: ViType employs the piezoelectric ceramic sensor to convert the vibration signals into recordable electrical signals which are then denoised using a filter and segmented by a double threshold-based mechanism.

2) Keystroke localization: Once the vibration signals are received and detected, ViType utilizes a keystroke localization algorithm to extract the unique vibration feature (i.e., power spectrum density) in the frequency domain and fuses it with amplitude signals as inputs to a trained classifier based on ANN for the purpose of localization.

3) Runtime calibration and adaptation: ViType takes the advantage of user's on-screen feedback to correct accidental classification errors and this process is called calibration. Furthermore, it adopts a runtime adaptation algorithm to update and enlarge the training set to maintain the high accuracy of classification.

Fig. 3 describe the work-flow of ViType system. In the initial training stage, the vibration signals of keystrokes are sensed, denoised and segmented by the detection mechanism. Afterwards, the localization algorithm extracts the feature and builds the ANN model. When a new keystroke is detected, the localization algorithm finds the best match in the trained model and then output the resulting number on the screen along with candidate keys which can be calibrated by the user. The feedback of calibration is transmitted to the runtime adaptation algorithm to update the training set.

IV. ViType

A. Keystroke Detection

1) Sensing: There are inertial measurement units (IMU) in the COTS wearable wristbands which are able to detect vibration. However, they are engineered for very different applications such as motion tracking rather than measuring acoustic signals propagated through the human body. Consequently, they are unfit in many crucial ways [3]. Piezoelectric ceramic sensor uses the piezoelectric effect to measure the vibration intensity by converting it to an electrical charge. In a piezoelectric ceramic device, mechanical stress, instead of externally applied voltage, causes the charge separation in the individual atoms of a material. Thus, the vibration caused by finger taps is able to be converted to an electrical charge. Fig. 2 shows a sample piezoelectric ceramic sensor whose external diameter is 20 mm and thickness is only 0.4 mm (FT-20T-6.5A1). The small form factor of the sensor makes it easy to be embedded to a smart wristband as a chassis.

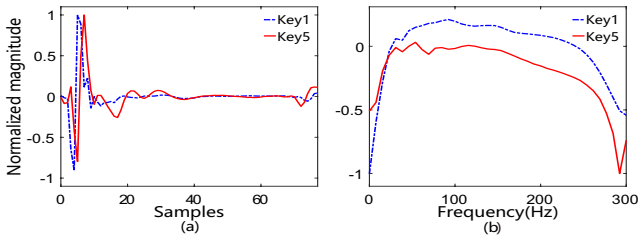

Fig. 4. Sample amplitude and PSD of key 1 and key 5.

2) Denoising: Unlike a microphone-based acoustic system, ViType is capable of resisting environmental noise using the vibration signal. Therefore, ViType has a low level of noise naturally. At first, we used a 20 Hz Butterworth high pass filter to remove the low frequency noise caused by the direct current component and the human mobility (less than 5 Hz). Second, we set the cut-off frequency to 300 Hz of a Butterworth low pass filter since the vibration signal tapped on the opisthenar is realized in low frequencies (less than 200 Hz) domain. Then, the noise in the higher frequency domain can be removed as well.

3) Segmentation: We use energy-based double threshold approach to detect the start point of a keystroke [19]. The lower threshold is $\mu+\sigma$ and the higher one is $\mu+3\sigma$, where μ and σ are the mean and standard deviation of energy obtained from collected signals, respectively. The lower one is very sensitive to the variation of signal and can easily be broken, whereas the higher one will not. Exceeding lower threshold level is not necessary to detect the start point as there might be some noise whose energy is higher than it. Only when the high threshold is overpassed, the low threshold can be considered as the start point of the signal to be detected. In terms of the end point, we set it at 0.1 s after the start point as the duration of a keystroke signal is usually around it.

B. Keystroke Localization

1) Feature Selection: The attenuation model of vibration signals, as stated by the relation in (1), provides us a hint of using raw data of amplitude as location signature and Fig. 4 shows the distinguishable vibration signals collected from key "1" and key "5". Furthermore, the signals also carry diminishing energy at different frequencies over different distances, and thus we investigate the profile pattern of key "1" and key "5" in frequency domain. Specifically, we choose PSD of the collected vibration signals, which reveals the power distribution in different frequency. If k_i is the received vibrations signals, then the PSD can be defined as

$$PSD_i = 10\log_{10}\frac{(abs(FFT(k_i))^2}{f_s \times n}, \tag{2}$$

where $FFT(\cdot)$ is the fast Fourier transform operation, f_s is the sampling rate, and n is the number of samples of

received signal k_i. Fig. 4 shows that the PSD features of two keys have different profile, which exhibits distinct values across frequencies. This give us another justification for using PSD to locate the keystrokes. Therefore, we decide to extract amplitude and PSD from raw signal and fuse them together as the inputs to estimate the classification model in ViType system. Note that the profile of each key shows the distinction but we only present the profiles of 2 keys for the brevity of image expression.

2) Classification Algorithm: An ANN is a computing system inspired by the biological neural networks that constitute animal brains. The back-propagation (BP) algorithm is one of the well known methods of ANN. In addition, unlike Deep Learning methods which have heavy computational requirement, BP ANN is suitable to apply in the smart wristband that has limited resource for computation. BP ANN is based on gradient descent method which minimizes the sum of the squared errors between the actual and the desired output values. The basic formula of BP algorithm [20] is

$$W(n) = W(n-1) - \Delta W(n), \tag{3}$$

where

$$\Delta W(n) = \eta \frac{\partial E}{\partial W}(n-1) + \alpha \Delta W(n-1), \tag{4}$$

where W, η, E, and α Δ W(n-1) are weight, learning rate, gradient of error function, weight incremental quantity, respectively.

Since we apply gradient descent method to find the optimal solution, it is necessary to normalize the data. Otherwise, it is difficult to converge. Moreover, after the normalization, the speed of finding optimal solution increases and hence the classification accuracy increases as well. Thus, we employ the min-max normalization concept shown in (5) to linearly transform the raw data x and make the resulting values to be within [0,1].

$$x' = \frac{x - min(x)}{max(x) - min(x)}. \tag{5}$$

One of the missions of BP ANN when finding the optimal model is to minimize the loss function. Since the keystroke recognition is a classification problem, thus we choose the typical cross-entropy loss function to train the BP ANN.

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i \log \hat{y}_i - (1-y_i)\log(1-\hat{y}_i) \tag{6}$$

where N is the number of input, y_i is the ground truth and $\hat{y}_i$ is the predicted output.

However, when the training set is limited like that in ViType (maximum of 70 for each key), finding minimum of $L(w)$ is not the best scheme because of the over-fitting phenomenon. Therefore, to avoid the over-fitting phenomenon and improve generalization, we add a penalty term $\frac{\lambda}{2}||w||^2$ to $L(w)$, which is termed as regularization. Thus, the new loss function after regularization is

$$Loss' = Loss + \frac{\lambda}{2}||w||^2 \tag{7}$$

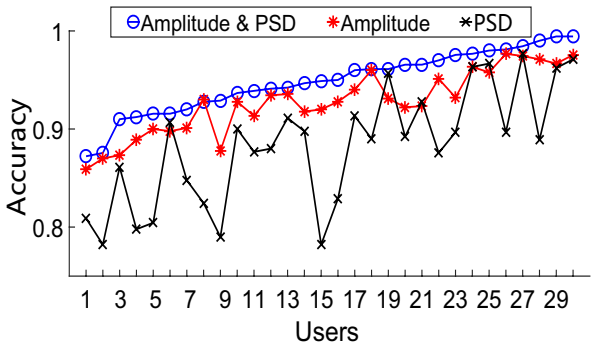

Fig. 5. Classification accuracy of three feature subsets.

The penalty term $\frac{\lambda}{2}||w||^2$ consists of the mean of the sum of squares of the network weights and biases.

In the application scenario of ViType, when a user wears our wristband for the first time, a short initial phase (tapping each key 20 times within 3 minutes) is needed for training an ANN model. After the training, each time the user wears the wristband, the previously learned model is used to recognize the keystrokes.

C. Runtime Calibration and Adaptation

ViType designs a runtime calibration and adaptation system to adapt itself to the deviation of wristband and keystrokes, and keeps the training data as new as possible to achieve better localization accuracy. As for calibration, for each keystroke, besides the output which is given by the classification algorithm, ViType also displays other top 2 candidate keys. A user can click any candidate key if it is the actual intended key when the algorithm gives a wrong output on the touch screen. If there is no intended key contained in the candidate list, the user will require to turn to the built-in on-screen keyboard.

In terms of practical usage, there are three cases: (1) a user does not select the candidate key and ViType will deem the localization output as correct, (2) a user selects any candidate key, which means that the system gives a wrong output and ViType maps the current input signal with candidate key rather than the wrong output, (3) a user taps the "Delete" button and it is not necessarily a hint for localization error as it may be the user's own input error. Therefore, for adaptation, we have designed a special scheme to update the training set. For case 1, the input sample will be added only once into the queue corresponding to the correct output. For case 2, the input sample will be added t_k times into the queue corresponding to the selected candidate key. Note that t_k is defined as consecutive error times of key k and varies from 1 to 4. For instance, if the system gives wrong output of key k consecutively for 3 times, the value of t_k will be equal to 3. If the wrong output of key k occurs consecutively more than 4 times, the maximum of t_k will be 4. Once the system produces the intended output for key k, t_k is reset to 1. We define the number of sample in queue of key k as Q_k. Then we have the total number of samples in all queues:

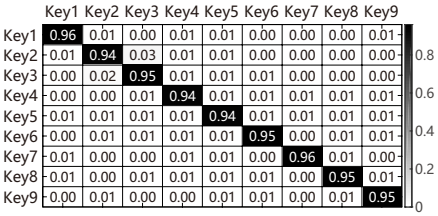

Fig. 6. Confusion matrix of 9 keys using amplitude & PSD as features.

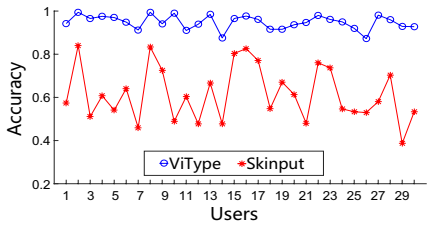

Fig. 7. Comparison of classification accuracy between ViType and Skinput.

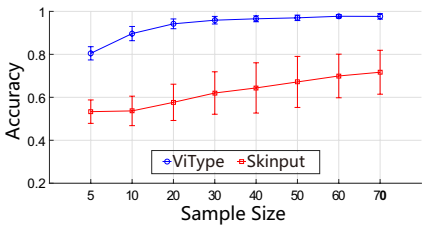

Fig. 8. Impact of initial training set size for ViType and Skinput.

$$N = \sum_{k=1}^{9} Q_k. \tag{8}$$

Once N is larger than 20, the samples in all queues will be popped into the training set and the ANN model is trained again. Note that the oldest samples leave the corresponding queue if the training set sizes reach the maximum of 70 for each key.

V. Implementation & Evaluation

Implementation: In this prototype, we implemented ViType using a piezoelectric ceramic sensor and an amplifier connected to a Raspberry Pi controller via an Analog to Digital Converter (ADC). The Body vibrations are collected via BCM2835 Library with C. Then, we transmit them to a conventional desktop computer by a PL2303 USB To Transistor-transistor logic (TTL) Converter Adapter Module. It is implemented via WiringPi Library with C.

As for the keystroke recognition, the signal denoising, keystroke detection and BP ANN are implemented in Matlab toolbox. Note that we set the number of hidden layers to 1 and the number of hidden nodes to 140 and the learning rate to 0.1 for BP ANN.

Experimental Setup: We recruited 30 participants (20 of them are male) who are in the age range of $[19-24]$ and stand for the crowd that are most likely to use our system. Besides, their body mass indexes (BMIs) range from 17.26 (lean) to 29.38 (obese). Note that all the experiments involving human subjects conformed to the relevant regulations of our university.

The evaluation experiments are launched in a conventional office environment. At the beginning of the experiments, the instructor marked the location of each key using a marker. This is because the participants may not realize the fact that tapping a bony area produces more stable vibration signal compared to the case when fleshy area of the back of one's hand is tapped [3]. Moreover, participants are given a 10-minute warm-up period to become familiar with our system before the experiments. In all experiments, we adopt the following default setting unless explicitly specified. The participants are instructed to tap 30 times on each key in an orderly fashion (270 examples for each person, 8100 data points in total). For example, we ask the participants to tap on key 1 for 30 times, then key 2 for 30 times, and so on. We then apply 10-fold cross-validation method to initialize the BP ANN learning scheme to estimate the mathematical model. Note that the calibration and adaptation scheme is turned on for the experiments in Section V-C only.

A. Accuracy of ViType

In this section, we first verify the suitability of the features we used by comparing three other feature subsets. Then, we validate the accuracy of ViType in terms of the keystrokes detection and keystrokes classification. Afterwards, we evaluate the accuracy comparing with the existed work Skinput [3]. We end this section with the discussion about the impact of training sets of different sizes on the keystrokes recognition accuracy.

1) Effect of feature subset: We have got two potential features for classification in Section IV-B, and hence we obtain three different feature subsets, which are (1) only amplitude, (2) only PSD, (3) fusion of amplitude and PSD. In this experiment, we investigate the classification accuracy with respect to different features mentioned above using BP ANN. As shown in Fig. 5, the feature set (3) obtains the highest average accuracy at 94.8%, followed by (1) at 92.7% and (2) at 88.3%, respectively. The amplitude of each keystroke reflects the attenuation information in time domain while the PSD reflects it in the frequency domain. Consequently, the fusion of these two features complements each other and improves the classification performance.

2) Baseline detection and localization: We perform the experiment of keystroke localization following the setting discussed above. After collecting the data, we obtain the results with 0% mis-detection and 0% flase-alarm. In terms of localization, Fig. 6 plots the resulting confusion matrix of localization accuracy for 30 participants, showing that the average classification accuracy is 94.8%.

3) Comparison with Skinput: Here, we compare the performance of ViType to the state-of-the-art approach Skinput, in which signals are collected from 10 piezo films of an armband at a very high sampling rate. Whereas, ViType uses a sensor with small size for making it easier and more cost effective to be embedded on smart wristbands. Moreover, ViType samples at an order-of-magnitude lower rates that makes it more efficient to run on resource limited smart wristbands. In this experiment, we input the extracted features adopted in Skinput of the same raw data into a SVM classifier (used in Skinput as

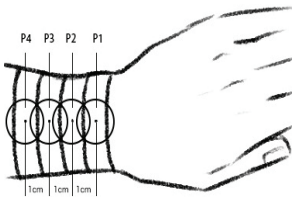

Fig. 9. Positional variation of wristband.

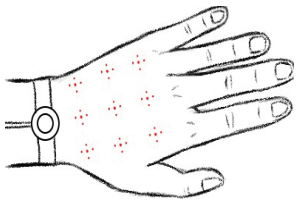

Fig. 10. Positional variation of taps.

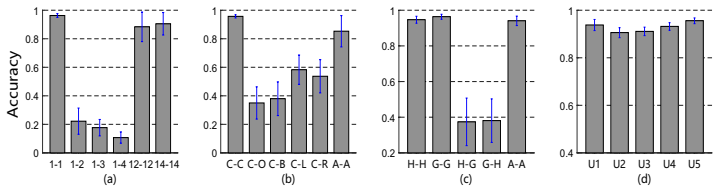

Fig. 11. (a) Accuracy of positional variation of wristband. (b) Accuracy of positional variation of taps. (c) Impact of different force taps. (d) Impact of tapping while in the walking phase.

well). Fig. 7 demonstrates how ViType outperforms Skinput. Comparing the red line with the blue line, we observe that Skinput obtains an average accuracy of 61.6%, while ViType can obtain an average accuracy of 94.8%. In other words, ViType can obtain an approximate 1.52 times higher average classification accuracy compared to Skinput. Furthermore, ViType has a relatively steady accuracy among different users while the accuracy of Skinput has a large differences among people and the standard deviation of accuracy are 0.032 and 0.124 for ViType and Skinput, respectively. We note that the worst case of ViType outperforms the best case of Skinput.

4) Impact of training set size: Intuitively, the classification accuracy of our system can be enhanced by enlarging the size of the training set. This is due to the fact that it is rather difficult for ViType users to tap exactly on the same point for each key without any deviation. To verify this hypothesis, 8 participants are asked to tap 80 times per key in order to produce a training set. We calculate the accuracy with respect to different size of training set (from 5 to 70) for both ViType and Skinput and plot Fig. 8.

We can observe evidently that the classification accuracy rises upward monotonically with the increasing size of the training set for both ViType and Skinput. However, the accuracy of ViType increases faster than that of Skinput when the size of training samples rises from 5 to 20 (at about 80% and 94%, respectively), while the accuracy of Skinput is below 60% in this range. This implies that ViType has a better user experience as we may ask a user to tap 20 times for each key only to initialize the training sets, the duration of which is within 3 minutes. ViType has the calibration and adaptation scheme. Hence, as a user types every day, its training set grows continuously (e.g., even 70 for each key), which means that the accuracy of ViType can approach nearly 98% (see Section V-C for more details).

B. Robustness of ViType

In this section, we focus on the robustness of the system and employ our system under several different conditions. Note that we did not turn on runtime calibration and adaptation feature of ViType in these experiments. The issues, in which we are concerned, are stated as follows.

- Positional variation of wristbands
- Positional deviation of tap
- Difference tap force
- Mobility

1) Positional variation of wristbands: The key insight of ViType is the distinctive vibration signal produced by tapping on different locations on the opisthenar which requires to fix the position of the wristband. However, there is quite a common scenario that the wristbands user wearing shifts from original location to others over the time of usage. In this experiment, we assume that the original position of the sensor is point P1. We then have points P2, P3 and P4 by shifting 1 cm towards the elbow gradually (Fig. 9). We change the location of the sensor from P1 to P4 and ask 8 participants to consecutively tap on each key for 30 times. In Fig. 11(a), the X-axis is the format of "training data-test data". For example, "1-2" indicates that we train the classifier with 20 samples collected on P1 and test the system with 10 samples collected on P2. Particularly, "12-12" means that we train the classifier with 20 samples from P1 and P2, respectively (40 training samples in total), and test the system with the remaining 10 samples from P1 and P2, respectively. We have 2 observations in the context of this experiment: (i) using the samples collected from the same location for the test purpose (e.g., 1-1) can achieve much higher accuracy comparing with 1-2, 1-3, 1-4 cases. (ii) initializing the system from the samples collected from different points (e.g., 12-12, 14-14) can mitigate the impact of wristband displacement. Consequently, it provides us a hint of designing a runtime adaptation scheme to update the training set (see Section V-C for the details).

2) Positional deviation of tap: Even for tapping on the same key, the slight deviation of each tap occurs all the time. To investigate the impact on the performance of deviation of taps, we ask 8 participants to consecutively tap on each key for 30 times as well as the deviation key which are illustrated in Fig. 10 (with 0.5 cm interval from the center of key in four directions, namely over (O), below (B), left (L) and right (R)). In Fig. 11(b), similar to Fig. 11(a), the X-axis is also the format of "training data-test data". For instance, C-O indicates that we train the classifier with 20 samples collected on center and test the system with 20 samples collected on the over location. Particularly, "A-A" means that we train the classifier via 100 samples collected from every point (20 from each location) and test the system with the rest of 50 samples (10 from each location). The histogram shows that when the training set and the test set are from different locations, the localization accuracy suffers a great drop to less than 60%. However, the

accuracy of the "A-A" case recovers to around 83%, which means that ViType has the resilience to tapping deviation when the training set is associated with the tapping deviation.

3) Force of Tap: The resultant vibration signal may be different when users apply different tapping force, which results in localization errors. To examine the impact of tapping force, we conduct this experiment in which we ask 8 participants to tap on each key 30 times both gently and heavily, that results in 4,320 responses (9 keys $\times 8$ users $\times 30$ times $\times 2$ ways). Note that "H" indicates that a user taps heavily while "G" indicates that a user taps gently. Fig. 11(c) shows the resulting graphics. Similar to the previous sections, the X-axis stands for "training force-testing force". For example, H-G indicates that we train our classifier with 20 samples collected when participants tap heavily and test with 10 samples data collected when participants tap gently. Notably, A-A means that we initialize the classifier with 20 samples and test with the remaining 10 samples (half gently and half heavily). From the figure, we discover that the classification accuracy drops to below 40% when the testing tap force is different from the training tap force (i.e., G-H and H-G). A different tap force incurs different characteristics of signals, which leads to lower classification accuracy. Moreover, while initializing our classifier with both "heavy" and "gentle" data, the accuracy recovers to the same level with the accuracy of using the same force (H-H and G-G).

In reality, users may apply different tap force to make the training set more approximate to the results in "A-A" model, which may contain all kinds of tap force in the test phase.

4) Mobility: If users have other physical movements while they type on ViType, it will probably result in noise interference and cause a higher detection error rate. This happens regularly in our daily life. For example, we may need to send an important message or chat with someone when we are in the walking phase. Practically, we cannot avoid noise interference due to this movement. To investigate how mobility impacts the classification accuracy, we conduct the following experiment to study the accuracy of our system while walking and typing simultaneously. In this experiment, we ask five of our participants to tap on nine keys 30 times each respectively when walking, and then apply 10-fold cross-validation technique to evaluate the accuracy. Note that we did not study how our system performs when users are jogging since it is too dangerous to type in this scenario and we strongly advise the users to avoid typing when jogging. Fig. 11(d) plots the results of this experiment, which shows the individual accuracy of every participant. Compared with participants sitting in an office, ViType still obtains a high accuracy (92.8% on average). The reason of high accuracy is that the noise caused by human mobility is at low frequency (less than 5 Hz) and we remove it via a 20 Hz Butterworth high pass filter.

C. Runtime calibration and adaptation

While using ViType in practice, it is difficult for users to keep the position of the wristband and keystroke unchanged while the usage period of time. It is quite unfriendly for users if they need to reinitialize the system every time they use ViType. ViType has proved its robustness and indicates that we can alleviate the deterioration of accuracy by increasing the training samples collected in different conditions. However, we still want to achieve better accuracy and provide better user experience in reality. Hence, we design the calibration and adaptation system. In the following two experiments, we turn on the runtime calibration and adaptation scheme.

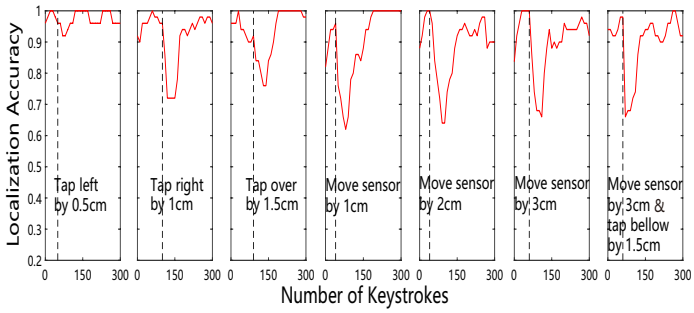

Fig. 12. Runtime calibration and adaptation scheme helps ViType to restore its high accuracy after deviation of watch and keystrokes (dotted lines denote the moment of displacement occurrence).

1) Resilience to displacement: In the first experiment, we count the localization accuracy averaged over the last 50 keystrokes with respect to the displacement of wristband and tap position in different levels. The resulting impact on accuracy is shown in Fig. 12. In terms of the displacement of the wristband, the localization accuracy drops to around 65% if the wristband is moved from the original position no matter with a displacement of 1 cm, 2 cm or 3 cm. As for the displacement of tap position, in the case with the smallest displacement at 0.5 cm, the accuracy shows no significant degradation. Whereas in the case with larger displacement of 1 cm and 1.5 cm, the accuracy drops to about 72% and 76%, respectively. Particularly, when the displacement of wristband and tap position occur at the same time, the accuracy drops to around 62%. However, ViType's calibration and adaptation scheme can mitigate the impact of these wristband and tap position displacement and recover the accuracy to above 95% after a few tens of inputs.

2) Temporal stability: In order to judge this metric, we conducted experiments 5 times over the interval of 1 hour, 1 day, 2 days, 1 week and 1 month. In each time, we tap from key "1" to key "9" for 100 rounds (900 keystrokes in total) while considering the average localization accuracy of the past 50 keystrokes. We observe that as the size of the training samples is enlarged, the localization accuracy remains stable at around 98% each time. This indicates that ViType is temporally stable over the time.

D. Cost

Firstly, users only need to input 20×9 training instances at the beginning (all 30 subjects finished tapping within 3 minutes of our evaluation). Then, the training duration of BP ANN is only 0.6 s. Moreover, we measure the latency between

each tap and ViType outputs (i.e., the localization result) on the screen. The results show that the classification latency is around 0.2 s, which is well below the human response time. Therefore, there is no lagging effect when users use ViType. Furthermore, ViType with a low sampling rate (e.g., 600Hz) is significantly more efficient to run on energy-limited smart wristbands compared with state-of-art approach Skinput. When it comes to the case of hardware expense, since we only employ one piezoelectric ceramic (at 0.15 dollar) and one amplifier (at 0.45 dollar), it is not expensive for a manufacturer to embed ViType on a smart watch.

VI. Conclusion

This paper presents a novel text input system for wristbands assuming the back of one's hand as a virtual keyboard. Body vibration is detected by a small sensor embedded to the wristbands at a lower sampling rate, and then classified by BP ANN to support text input. ViType achieves high keystrokes recognition accuracy and is also robust under several realistic text input conditions such as tapping with a different force, typing when walking and so on. The result demonstrates that a neural network works well for classifying body vibration, and ViType is more accurate and robust with more training samples collected under different conditions and the update of a training set can be done by the calibration and adaptation feature of ViType.

VII. Acknowledgement

This research was supported in part by the China NSFC Grant 61472259, 61502313, Joint Key Project of the National Natural Science Foundation of China (Grant No.U1736207), Guangdong Natural Science Foundation 2017A030312008, Shenzhen Science and Technology Foundation (No. JCYJ20170302140946299, JCYJ20170412110753954), Fok Ying-Tong Education Foundation for Young Teachers in the Higher Education Institutions of China(Grant No.161064), Guangdong Talent Project 2014TQ01X238, 2015TX01X111 and GDUPS (2015), Tencent Rhinoceros Birds - Scientific Research Foundation for Young Teachers of Shenzhen University, SZU R/D under Grant 2016044, Lu WANG is the corresponding author.

References

[1] R. Nandakumar, V. Iyer, D. Tan, and S.Gollakota (2016, May). Fingerio: Using active sonar for fine-grained finger tracking. In Proc. ACM CHI (pp. 1515-1525). ACM.

[2] W. Wang, A. X. Liu, and K. Sun (2016, October). Device-free gesture tracking using acoustic signals. In Proc. ACM MobiCom (pp. 82-94). ACM.

[3] C. Harrison, D. Tan, and D. Morris (2010, April). Skinput: appropriating the body as an input surface. In Proc. ACM CHI (pp. 453-462).

[4] A. Abdullah and E. F. Sichani (2009). Experimental study of attenuation coefficient of ultrasonic waves in concrete and plaster. The International Journal of Advanced Manufacturing Technology, vol. 44, no. 5-6, pp.421-427.

[5] V.Lakshmipathy, C.Schmandt, and N. Marmasse (2003, November). TalkBack: a conversational answering machine. In Proc. ACM UIST (pp. 41-50).

[6] J. Wang, K. Zhao, X. Zhang, and C. Peng (2014, June). Ubiquitous keyboard for small mobile devices: harnessing multipath fading for fine-grained keystroke localization. In Proc. ACM MobiSys (pp. 14-27).

[7] C. Harrison, H. Benko, and A. D.Wilson (2011, October). OmniTouch: wearable multitouch interaction everywhere. In Proc. ACM UIST (pp. 441-450).

[8] M. Ogata, Y. Sugiura, H. Osawa and M. Imai (2012, October). iRing: Intelligent Ring Using Infrared Reflection. In Proc. ACM UIST (pp. 131-136).

[9] S. Nirjon, J. Gummeson, D. Gelb, and K. H. Kim (2015, May). Typingring: A wearable ring platform for text input. In Proc. ACM MobiSys (pp. 227-239).

[10] W. Kienzle and K. Hinckley (2014, October). LightRing: Always-available 2D Input on Any Surface. In Proc. ACM UIST (pp.157-160).

[11] M. Prtorius, A. Scherzinger, and K. Hinrichs (2015, September). SkInteract: An on-body interaction system based on skin-texture recognition. Springer Human-Computer Interaction (pp. 425-432).

[12] T. Sekitani, M. Kaltenbrunner, T. Yokota, and T. Someya (2014, June). Imperceptible Electronic Skin. In Proc. SID Symposium Digest of Technical Papers (Vol. 45, No. 1, pp. 122-125).

[13] G. Laput, R. Xiao, X. A. Chen, S. E. Hudson, and C. Harrison (2014, October). Skin buttons: cheap, small, low-powered and clickable fixed-icon laser projectors. In Proc. ACM UIST (pp. 389-394).

[14] S. Nirjon, J. Gummeson, D. Gelb, and K. H. Kim (2015, May). Typingring: A wearable ring platform for text input. In Proc. ACM MobiSys (pp. 227-239).

[15] H. Wen, J. Ramos Rojas, and A. K. Dey (2016, May). Serendipity: Finger gesture recognition using an off-the-shelf smartwatch. In Proc. ACM CHI (pp. 3847-3851).

[16] R. Xiao, G. Lew, J. Marsanico, D. Hariharan, S. Hudson, and C. Harrison (2014, September). Toffee: enabling ad hoc, around-device interaction with acoustic time-of-arrival correlation. In Proc. ACM MobileHCI (pp. 67-76).

[17] S.Pan, C. G. Ramirez, M. Mirshekari, J. Fagert, A. J. Chung, C. C. Hu, J. P. Shen, H. Y. Noh and P. Zhang (2017, April). SurfaceVibe: vibration-based tap and swipe tracking on ubiquitous surfaces. In Proc. ACM IPSN (pp. 197-208).

[18] J. Liu, Y. Chen, M. Gruteser and Y. Wang (2017, June). VibSense: Sensing Touches on Ubiquitous Surfaces through Vibration. In Proc. IEEE SECON (pp. 1-9).

[19] W Chen, M Guan, L Wang, R Ruby, K Wu.(2017, July). FLoc: Device-free passive indoor localization in complex environments. In Proc. IEEE ICC (pp. 1-6).

[20] W. Jin, Z. J. Li, L. S. Wei, and H. Zhen (2000). The improvements of BP neural network learning algorithm. In Proc. IEEE WCCC-ICSP (Vol. 3, pp. 1647-1649).

[21] W. Chen, Y. Lian, L. Wang, R. Ruby, W. Hu, and K. Wu (2017, November). Demo: Virtual Keyboard for Wearable Wristbands. In Proc. ACM Sensys (pp. 44-45).

[22] J. Wang, R. Ruby, L. Wang, and , K. Wu (2016, December). Accurate Combined Keystrokes Detection Using Acoustic Signals. In Proc. IEEE MSN (pp. 9-14).

[23] Y. Wang, K. Wu and L. Ni (2017). WiFall: Device-free Fall Detection by Wireless Networks. In IEEE Transactions on Mobile Computing, Vol. 16, No. 2, (pp. 581-594).

[24] Y. Zou, J. Xiao, K. Wu, J. Han, Y. Li and L. Ni (2017). GRfid: A Device-Free RFID-Based Gesture Recognition System. In IEEE Transactions on Mobile Computing, Vol. 16, No. 2, (pp. 381-393).

[25] L. Wang, X. Qi, J. Xiao, K. Wu, M. Hamdi and Q. Zhang (2016). Exploring Smart Pilot for Wireless Rate Adaptation. In IEEE Transactions on Wireless Communications (pp. 4571-4582).

[26] Y. Tong, L. Chen, Z. Zhou, H. Jagadish, L. Shou, W. Lv (2018). SLADE: a smart large-scale task decomposer in crowdsourcing. To appear in IEEE Transactions on Knowledge and Data Engineering.

[27] G. Wang, Y. Zou, Z. Zhou, K. Wu and L. Ni (2016). We Can Hear You with WiFi. In IEEE Transactions on Mobile Computing, Vol. 15, No. 11 (pp. 2907-2920).

[28] Y. Zou, W. Liu, K. Wu and L. Ni (2017, October). Wi-Fi Radar: Recognizing Human Behavior with Commodity Wi-Fi. In IEEE Communications Magazine, Volume: 55, Issue: 10 (pp. 105-111).