2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)

Oinput: a Bone-Conductive QWERTY Keyboard Recognition for Wearable Device

Yongzhi Huang, Shaotian Cai, Lu Wang, Kaishun Wu
College of Computer Science and Software Engineering, Shenzhen University

***Abstract*—The emergence of wearable devices has brought great simplicity and convenience to people's daily lives. However, due to the small form-factor, low-profile hardware interfaces, the input scheme for such wearable devices becomes a bottleneck and even sabotages their functionalities. The state-of-the-art interaction schemes, including voice input, inertial measurement unit (IMU) based input, or acoustic based input, all require a stable environment, which is critical for wearable device. To break this stalemate, we propose a stable QWERTY keyboard input for wearable devices based on bone-conduction models. Using the characteristics of human anatomy, we achieve a low-cost and high precision text input system, named Osteoacusis input (Oinput), with the help of human bones. To be specific, we first investigate a new set of bone-conduction theories. Through this set of theories, we combine a strong anti-noise cyclic neural network to achieve a high-precision QWERTY keyboard recognition for text input. Furthermore, in order to improve the user experience, we leverage slightly keyboard layout changing, dimensionality and feature selection to reduce the power consumption while preserving the convenience and stability. We have conducted experiments on 30 volunteers. The results show that Oinput has superior robustness with a high recognition accuracy of 93.3% in average. Moreover, Oinput's calibration mechanism increases the accuracy by more than 99%.**

I. Introduction

Recently years, smart devices have boosted and played an nontrivial role in facilitating our daily lives. Intelligent devices are designed to be smaller and smaller to make them more portable, such as smart watches, glasses, bracelets and so on. The small form-factor, low-profile hardware interfaces in these portable smart devices make the user interaction experience extremely poor, and even sabotages their functionalities.

As the screen is too small to complete the text input in wearable devices, voice input has received a lot of popularity. Unfortunately, voice input highly depends on the user's accent, speech rate, and the network environment. As long as one of these three conditions is not met, the input accuracy is highly degraded. Besides, voice input is fragile to external noise and lack of privacy protection.

To fight against environmental noise and improve the interaction efficiency, acoustic based approaches emerged, such as FingerIO[1] and LLAP[2], which leveraged ultrasound to interact with smart devices by tracking user's fingers. Similarly, WristWhirl [3] and WatchWriter [4] harnessed in-built sensors, such as piezoelectric sensors, distance sensors, etc., to identify the user's gestures. However, gesture input is slow, and sometimes troublesome when the environment is not stable.

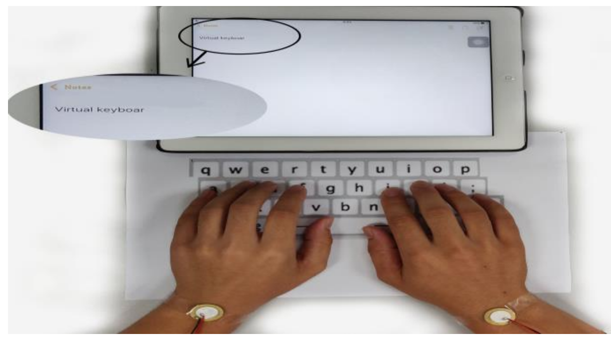


Fig. 1. A typical use case of Oinput.

A decent input system should satisfy the following three characteristics, stability, efficiency, and low cost. Stability means that the system is not easily affected by other external variables, such as environmental noise, scene changes or environmental factors. While Efficiency means that the system does not need complicated operations, such as extra training or tedious settings. Finally, the system should be low cost and easy to carry.

In this paper, we find that human organs have very strong anti-noise ability, and they maintain a state of dynamic balance throughout the year. Therefore, if we can rely on bones as a media for the typing system, the typing system satisfies the stability. In order to make the device low cost, we use a small piezoelectric ceramic vibration sensor with a coin-size. Finally, we design a QWERTY keyboard for an efficient input.

The idea is straightforward, yet there remain several challenges for implementation. First, what kind of features do we need to achieve a precise recognition. Second, as we only have ten fingers, and the number of buttons on a QWERTY keyboard are far beyond, how to realize a QWERTY keyboard with only ten fingers. Finally, due to the limited computational resources and battery life, how to design a practical input system with minimum energy consumption?

To overcome the above challenges, we propose an osteoacusis based QWERTY keyboard, namely Oinput, as shown in Figure 1. It harnesses only two coin-sized vibration sensor patches to help users realize a QWERTY -like keyboard on any plane. The vibration sensor on the wrists are used to collect the vibration signal generated by fingers tapping. Though

978-1-5386-7308-9/18/$31.00 ©2018 IEEE
DOI 10.1109/ICPADS.2018.00126

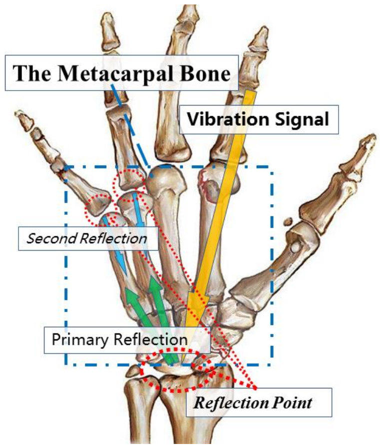


Fig. 2. Multipath effect in bone.

anatomical features, we analyze the vibration characteristics at different keystroke positions. We adopt a neural network with a very small amount of vibration signals for training (only about 5-10 training samples per button position) in less than 2 seconds. A user only needs to input according to the layout of the QWERTY keyboard, and the system completes the recognition of the input content. We have conducted extensive experiments on 30 volunteers. The input accuracy is as high as 93.3% in average.

The rest part of the paper is as follows. Section II introduces some background knowledge and the related work in this field. Section III describes the system architecture overview. Section IV introduces the detailed design of each module. In Section V, we evaluate the performance of the system through extensive experiments. Finally, we summarize the paper in Section VI.

II. Related Work & Motivation

A. Related Work

Today, portable devices only use voice input as a means of text input [15]. However, there are many problems with voice input, which we have discussed in the previous section. Mainly because it has higher requirements for users and networks. In addition, the voice input cannot be used when the outside sound is too loud or too small. If the external environment is too loud, the system cannot distinguish what the user says from the noise. When the outside voice is too whisper, such as in a library, it will bring a lot of embarrassment, and the use of it in public will also expose your privacy.

In order to get rid of the troubles caused by voice input, people try to use other ways to achieve input. FingerIO[1] and LLAP[2] use ultrasound that is inaudible to the human ear as a carrier. Use geometric methods to achieve precise positioning and tracking of the fingertips at the millimeter level. And through the depiction of a human finger trajectory, reaching recognition input. WristWhirl[3] and WatchWriter[4] complete the input by adding sensors such as piezoelectric sensors and distance sensors to the device to recognize the gesture of the user's palm. This method is very convenient to use for irregular interaction with the device. However, if it is used as input, it will give the user a very bad experience. Although the recognition of the handwriting input method has been completed as early as the last century, the use of handwriting for text input is very tiring and the efficiency is quite low.

TypingRing [11] designed a sensor similar to the ring. This ring carries the Inertial Measurement Unit (IMU). The user brings this ring to the middle finger. When the user needs to input, he only needs to control the middle finger, and select an area in the QWERTY keyboard that is divided into several areas (all composed of three buttons). Then tap one of the three fingers (index, middle, and ring finger) to select the three buttons in the area. This method of selecting regions is very time consuming, so it is difficult to complete fast input.

Wang [5] and Liu [6] used the method of keystroke positioning to identify the location of the keyboard. They all take advantage of some of the characteristics of sound signals, such as multipath overlays and Mel Frequency Cepstrum Coefficient (MFCC). However, when the objects of the environment (such as people) are constantly moving, the superposition of the sound multipath is constantly changing, so the recognition at this time becomes very difficult. Like voice input, MFCC is vulnerable to environmental interference and cannot be recognized properly.

MagBoard [13] uses the phone's magnetic sensor to identify the keyboard of its own design. This method can be used to design the keyboard according to your needs. However, when inputting, it is necessary to rely on magnetic things to stay on the keyboard of your own design, in order to be able to select the content to be input. If the dwell time is designed to be too short, the user needs to be hesitant to make a quick input, otherwise it is prone to false positives, resulting in poor user experience. Therefore, the input speed of this system is severely limited. At the same time, it is also very troublesome to carry magnetic equipment with you. The portable wearable smart device itself is very small, so the battery is very limited, however, the use of cameras [7] is very energy intensive.

Both Skinput[9] and ViType[10] are signals obtained by vibration sensors. Both systems need to tap some fixed places to complete the recognition of the input content by classifying the vibration signals. However, Skinput requires 10 sensors on an wristband to collect signals at very high sample rates (for example, 55 kHz). ViType [10] is even better, it uses a lower frequency (for example, 600Hz) to complete the identification of the opponent's back numeric keypad. Unfortunately, ViType requires the user to mark the 9 points of the numeric keypad on the hand, which is very unattractive. If the user do not want to mark a point on your hand, ViType requires the user to tap dozens of times around all possible points around each point. The user only places the center position of the 9 buttons and the eight directions around the center point (up, down, left, right, top left, top right, bottom right, and sit down). Tap 30 times for training, then the 9 keys need to be tapped 30*9*9=2430 times. Not only is the training set very large, but the entry into the training set is also very cumbersome.

Therefore, our system needs to meet three characteristics: stability, efficiency, and low cost. First of all, we want to design a system that does not need to carry overly complex equipment and does not require annoying preparation before use to achieve convenience. Second, we need the system to have a certain stability, and the recognition accuracy will not be greatly reduced

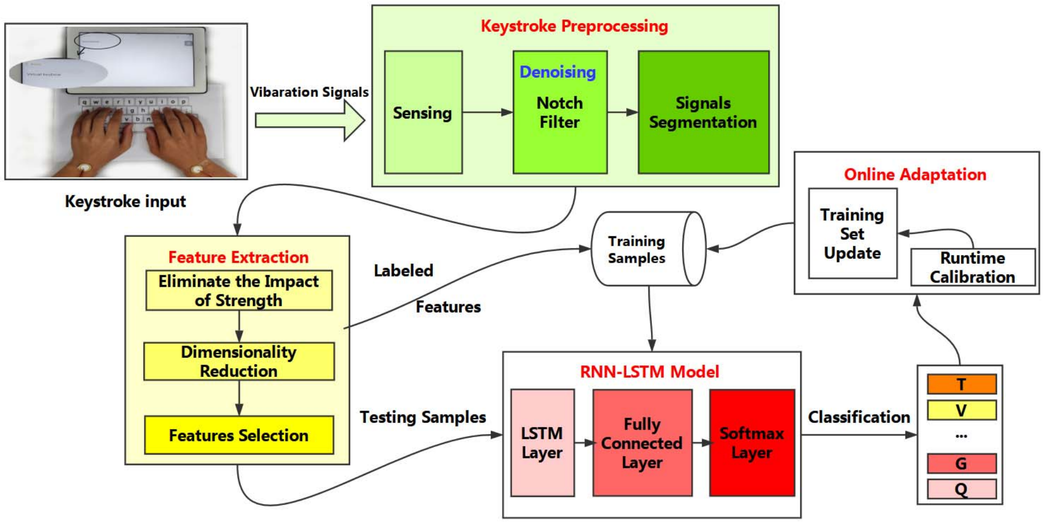


Fig. 3. Architecture of Oinput.

due to the noise of the environment, the movement of environmental objects or the transformation of the scene. Last but not least, we need the system to have low cost. Low cost refers not only to the cost of equipment, but also to the cost of power consumed by electricity, the time cost of typing, and the incalculable cost of loss due to privacy breaches.

B. Motivation

We have found through experiments that when people tap the keyboard, their fingers will generate vibration signals. The propagation of stress waves generated by free vibration in solids is relatively stable. In addition, the vibration sensor that detects the stress wave in the solid is a sensor with a lower price and a smaller volume in the sensor. What's more, it is convenient to use and the cost is low. Therefore, if the system can take advantage of the stress waves generated by the free vibration of the bone, the system will be able to meet convenience, efficiency and low cost.

Free vibration: Free vibration means that when an object is subjected to a force, if the force energy is strong enough, the surface of the object will be deformed. Deformation causes the object to vibrate freely. When the object is free to vibrate, its displacement changes sinusoidally with time, also known as hwristonic vibration. The amplitude and initial phase of the hwristonic vibration are related to the initial conditions. In other words, the ultimate vibration of the deformed object is related to the intrinsic properties of the object [20] (e.g., density, mass, and size). The frequency of this vibration is called the natural frequency of the object. For the n-order natural frequency, it can be expressed as

$$f_n = ck_n^{\ 2}W\sqrt{\frac{E}{\rho}(1-\xi^2)} \tag{1}$$

Where c is a constant term, k_n is related to the constraint of the object at the nth order, ρ is the density of the object, W is the thickness of the object, E is the elastic modulus, and ζ is the damping ratio. However, the wave equation generated by the propagation of vibration is expressed as

$$\frac{\partial^2(R\psi)}{\partial r^2} = \frac{1}{c^2}\frac{\partial(R\psi)}{\partial t^2} \tag{2}$$

ψ indicates that the point source from the sound source in the spherical wave surface range is the sound pressure at the r position, and R is the spherical distance from the point source of the spherical wave. C is the wave velocity, and the wave velocity is related to the elastic modulus and density of the medium, which can be expressed as $c = \sqrt{\frac{E}{\rho}}$. When the medium is a non-solid medium, E is the variable elastic modulus. When the medium is a solid medium, E is the shear elastic modulus or Young's modulus of the medium. Waves will form attenuation during the propagation process, and the attenuation coefficient α can be expressed as

$$\alpha^2 = \frac{\omega^2}{2c}\left(\frac{1}{\omega^2\tau^2} - 1\right) \tag{3}$$

Where ω is the circular frequency and τ is the relaxation time of the medium.

Propagation model in the human body: It is well known that in order to maintain the normal activities of life, mammals need the internal mechanisms of the body to create a stable environment for the cells. Such as acid-base balance, temperature and so on. We have found that the human body does have very strong anti-noise organs, such as bones. Since the bone density in the bone is always very stable, the natural frequency of the bone is generally constant according to the nature of the object's natural frequency [16, 17]. Therefore, we will build the system on the free vibration model of the bone. After research [16-19], the natural frequency of bones is around 200Hz. According to Shannon's law, the sampling rate needs to be at least twice the bandwidth. Therefore, we predict that the system sampling rate is around 400Hz, which is the lowest cost.

Since bones have such excellent characteristics, where can bones achieve normal input? We found that the hand bone is a good choice. Since different phalanx lengths are different,

Fig. 4. Oinput's unique keyboard, when detecting user input left and right, is considered to switch to digital input method.

different vibration signals can be generated. We found that vibration signals are reflected when the vibration signal propagates from the phalanx to the bones such as the scaphotrapeziotrapezoid at the bottom of the hand bone. These reflected vibration signals enter the metacarpal bone, creating a secondary reflection at the junction of the metacarpal and phalanx. Since the length of the metacarpal bone is also different, the time for all secondary reflected waves to reach the bottom of the palm is different, which results in a superposition of multipath signals. The vibration signals generated at different fingertips have different reflection angles at the bottom of the hand bone, so the final superimposed signals are also different. Using this principle, when using different fingers for tapping, the vibration signals collected by the sensor are different, as shown in Figure 2. In behavior, many unconscious movements are derived from muscle memory. That is to say, when using muscle memory, the same movement of people is almost exactly the same. Moreover, the relative position of the keyboard is fixed, so the shape of the finger joints is very similar when people tap the same button position on the keyboard. When the finger taps on different position of the keyboard, a vibration signal with a completely different propagation path can be generated. Method of maximum likelihood estimation [8], when the same button is tapped, the phenomena we can see is that a very similar vibration signal is generated.

III. System Overview

A. Design Goals and Challeges

In order to reduce the inconvenience caused by using too many sensors, we decided to use a piezoelectric ceramic vibration sensor to capture the vibration signal generated from the finger is tapped. Therefore, we placed a sensor on each of the two wrists to capture the vibration signal generated when the finger tapped the different button positions. However, the idea of implementing Oinput is difficult and needs to overcome the following challenges.

- In theory, the vibration signals at the same button position are the same, but in practice, even the signals at the same buttons are not exactly the same. When the user taps the desktop, it does not deliberately ensure that the strength of the tap is constant. The similarity is also very high when a finger is striking two adjacent button positions on the same line. Therefore, it is extremely challenging for precise keystroke recognition on a QWERTY keyboard layout.
- If we use complex algorithms and very high sampling rates, we can improve the recognition accuracy. However, due to the limited power of wearable devices, we need to strike a balance between the identification accuracy and energy consumption.
- To ensure that an input device works, it is important to be able to correctly identify what is being entered. However, if the system needs to spend a long time in identification, it will also make the user experience worse. Therefore, it is very important to be able to give users results efficiently.

B. Oinput System Overview

The system design of Oinput is shown in Figure 3. The Oinput system includes a piezoelectric ceramic vibration sensor. This device is capable of detecting very small vibration signals. Before amplifying the signal, we must use a notch filter to eliminate the electrical noise and prevent the effects of the gain. The noise-canceling signal is then amplified by an analog amplifier. Finally, these signal vibration signals are converted into digital signals by an analog to digital converter and sent to the host.

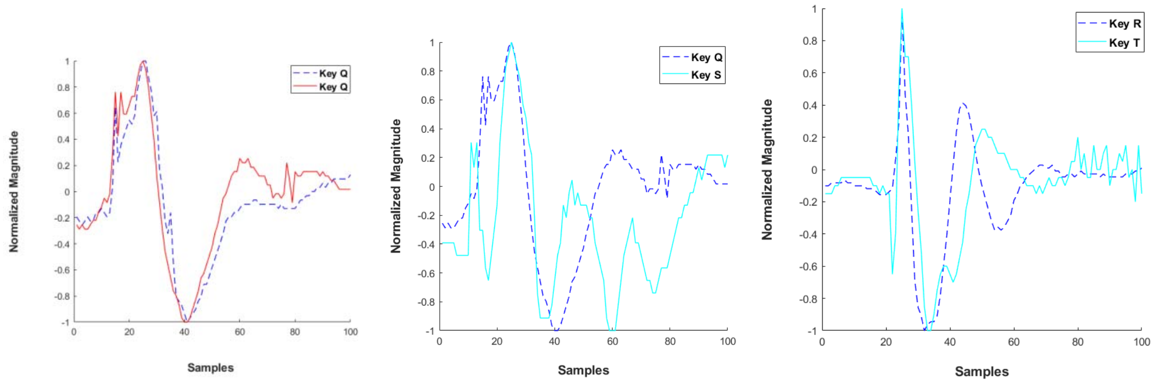
After receiving the data from the sensor, the receiving device processes and classifies the data and predicts the key content of the tap. Therefore, Oinput's software architecture is mainly divided into three parts, (1) Data preprocessing. (2) Key content recognition.

- **Data preprocessing:** The notch and bandpass filters are first used to filter out ambient noise and then the signal is cut by a dual threshold mechanism.
- **Identification of the key content:** Then, the system uses the algorithm to normalize the preprocessed signal, and the signal is in the same dimension, which can remove the difference of the signals generated by tapping the same key with different strengths to some extent. The next step is to reduce the dimensions. In this way, the amount of data that the system needs to process is reduced to improve the efficiency of the system. Not only that, but we also use feature screening to further reduce the amount of data in the system and improve the accuracy of recognition. After completing these two steps, the system enters the filtered features into the trained Recurrent Neural Network (RNN). The use of a cyclic neural network to share features in the temporal structure can solve the problem of vector limitation caused by the use of a Artificial Neural Network (ANN), thereby improving accuracy. The Long Short Term Memory (LSTM) algorithm is then used to solve the inefficiency caused by the gradient explosion. Finally, the network automatically gives the most likely results based on the characteristics of the input.

IV. System Design

A. Oinput's Keyboard

Oinput is a system that uses people's usual typing habits to identify keys. If you follow the typing habits, then each finger needs to map 4-8 keys, which is a very large workload. The more classifications that need to be made, the higher the accuracy of the data required. This may require an increase in the



(a) (b) (c)

Fig. 5. According to the input habit, tap: (a) two identical button positions (Key Q) on one plane, but use different strengths; (b) two different button positions (Key Q and Key S); (c) (using the index finger) two different button positions of the line (Key R and Key T); the generated vibration signal.

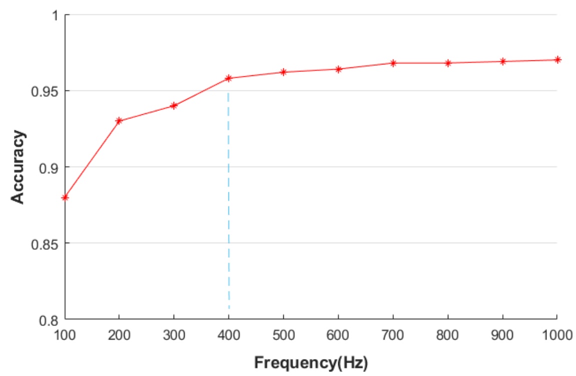


Fig. 6. Each button is tapped 30 times, and 5 samples are randomly taken as the training set, and the average recognition rate at different frequencies.

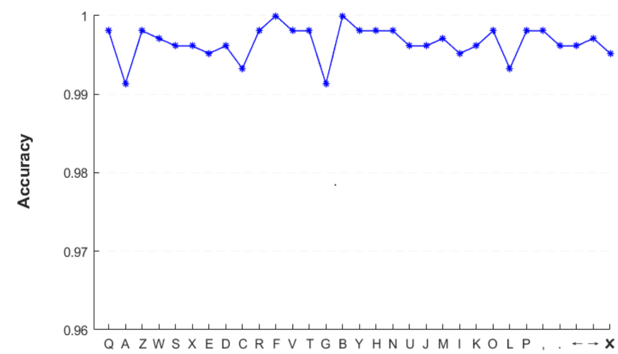


Fig. 7. Thirty volunteers used Oinput to tap each of the 31 buttons 100 times. Each button randomly took 30 samples as the training set, and the average recognition rate of the test set.

user's training sample data, or an increase in the sampling rate to be achieved. In order to make the device as energy-efficient as possible and do not want to increase the training burden on the user, we hope to start from other aspects. Through observation, we found that there are actually many buttons in the keyboard that people can't use when using portable smart devices. So, we decided to modify the keyboard, as shown in Figure 4. We removed the unused buttons and left 26 letters, semicolons, commas and question marks. In this way, each finger only needs to correspond to 3 or 6 buttons, which greatly reduces the burden on the system. We also divide the space key responsible for the thumb into three functions: left, right and space, so that the user can find the correct output by right and left selection in case of system identification error. Oinput also enables the ability to switch keyboards. When the user enters 'Left, Right, Right, Left', the system will switch to the numeric keypad. The numeric keypad has one finger for each button. If the user hits the same beat again with two thumbs, which is now '5, 6, 6, 5', it will switch back to the alphabetic keyboard.

B. *Eliminate the Impact of Strength*

Strength has always been one of the factors that influence recognition. ViType [10] requires users to distinguish between tap heavily and tap gently during training. Obviously, this is a very unwise way. The user is not a robot, and there are not only two options for tapping. Fortunately, we are using the inherent properties of bones. We found that even when we hit the power, the waveforms were still very similar. Therefore, we think of using the normalized method to put the signal into the same dimension, the expression is as follows,

$$y = \frac{(y_{\max} - y_{\min}) \times (x - x_{\min})}{(x_{\max} - x_{\min})} + y_{\min} \tag{5}$$

Where x and y correspond to the data before and after normalization. The result is shown in Figure 5(a). At this time, we found that the similarity of the two signals is very high. Through experiments, we can also find that, as shown in Figure 5, different buttons are different.

C. *Dimensionality & Energy Consumption Reduction*

As we mentioned earlier, the power of portable smart devices is limited. Therefore, we try to find ways to reduce the energy consumption of equipment. Because Oinput is divided into two steps, collecting data and predictions. In the input step, we can reduce the energy consumption by reducing the sampling rate. In the forecasting step, reducing the amount of calculation is the key to saving energy. After we achieved higher precision recognition, Oinput found the best sampling rate through more experiments. As shown in Figure 6, we let 30 volunteers tap each button 30 times and randomly take 5 samples as the training set to calculate the average recognition rate of the test set at different sampling rates. We found that the slope decreased when it was at 400 Hz, and the accuracy improvement was not obvious as the sampling rate increased. Therefore, we think that

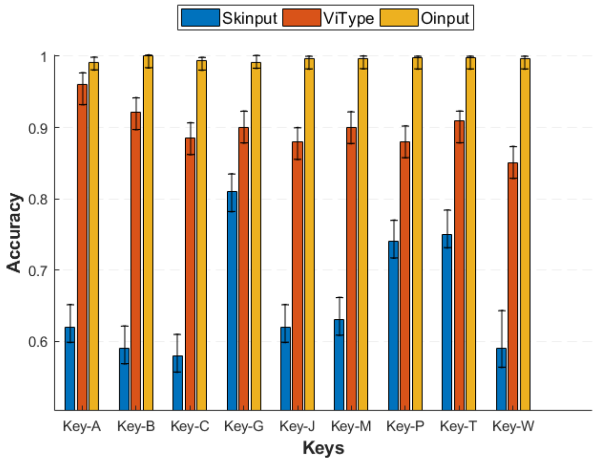


Fig. 8. Thirty volunteers used the trained system to achieve an average recognition accuracy of 100 single-letter inputs.

400Hz is a relatively energy-saving frequency, which corresponds to the natural frequency of bones mentioned above.

Therefore, we focus our attention on reducing the amount of calculations. Oinput uses the multipath effect because the paths of the different buttons are very different, which makes the reflected signals different in arrival time. We observed through the observation that the signals of different buttons have their own characteristics, and the features will not coincide with the signals of other buttons in the continuous sampling points. With this feature, we can reduce the dimensionality of the signal. After searching, we found that Haar Wavelet is a very good method. Haar Wavelet uses a very small amount of computation to divide the original signal into a trend sub-signal and a wobble sub-signal. The trend sub-signal can be expressed as follows,

$$a_m = \frac{f_{2m-1} + f_{2m}}{\sqrt{2}} \tag{6}$$

Among them, f_m is the sampling point of the original signal, and a_m is the trend sub-signal. Therefore, by using the trend sub-signal, not only the contour of the signal is maintained, but also the entire signal sampling point is reduced by half to achieve dimensionality reduction.

D. Feature Selection

In the previous section we mentioned that the signal is completely different after being affected by the multipath effect. On this basis, we believe that all the data of a signal is not useful, and the poor correlation with the label is a kind of interference to the neural network model. Therefore, we have designed a method of feature screening. η is the threshold, F is the feature, E is the mean, and the feature screening method can be expressed as follows.

$$T_{mi} = \begin{cases} 1 & (f_{mi} - E_f)^2 > \eta \\ 0 & else \end{cases} \tag{7}$$

$$F_m = \begin{cases} 0 & \sum_{i=1}^{n} T_{mi} > F_\eta \\ F_m & else \end{cases} \tag{8}$$

Using the feature screening method, you can remove useless signal features, eliminate interference, improve accuracy, and contribute to reducing energy consumption.

E. Neural Networks

In the process of selecting a neural network, we compared the ANN and the RNN separately. In the experiment, we found that the accuracy of the ANN recognition is very low if there are interference factors such as trains or subways. Our network uses LSTM (Long Short-term Memory). The LSTM network is a special kind of RNN that can learn long-term dependency information. The feedforward neural network has no persistent memory for sequence data, while LSTM can learn the long-term dependence of time series data. Also, it uses the accumulated form to calculate the state, which helps to solve the problem of gradient disappearance during training. So, it is often used for sequence data classification. In the LSTM network architecture, the network consists of five parts: the sequence input layer, the LSTM layer, the fully connected layer, the softmax layer, and the classified output layer.

V. IMPLEMENTATION & EVALUATION

A. Experimental setup

We recruited 30 volunteers between the ages of 19 and 22 (15 males and 15 females). These volunteers are able to complete the input without looking at the keyboard. All experiments involving human subjects are in accordance with the relevant regulations of our school. The evaluation of the experiment was carried out in a traditional office environment with a paper keyboard on the desktop. The volunteer was instructed to tap the button in an orderly manner. Each set of experiments requires 10 strokes for each of the 31 buttons, and a total of 310 strokes. After completing a set of experiments, rest for 2 minutes before proceeding to the next set of experiments. Both Skinput and ViType draw 9 numeric keypads on the back of the hand, requiring volunteers to control the strength of the tap when they are in the user's two systems. During the experiment, if there is no special explanation, Skinput and ViType mark the position of the tap on the back of the hand. In the experiment, if the user is required to use Skinput or ViType to specify a word, such as the letter B, then the user is required to tap the number key 2 twice, and the two taps can only be regarded as the result of one test. If there is no identification, it is required that the volunteers do not have a paper keyboard on the desktop when using Oinput, and only tap on the desktop by typing habits. When using ViType, volunteers are required to maintain the strength of the tapping, erase the position of the button marked on the back of the hand, and simply tap on the back of the hand.

B. Oinput System Performance

First we check the accuracy of the recognition of the Oinput system. It is mainly divided into three steps, first with a given sample size as the training set. The second step is to compare the number of samples with other systems to the training set, and compare the accuracy of the input letters to verify the practicability of the system. The third step is to use 1/3 of the sample

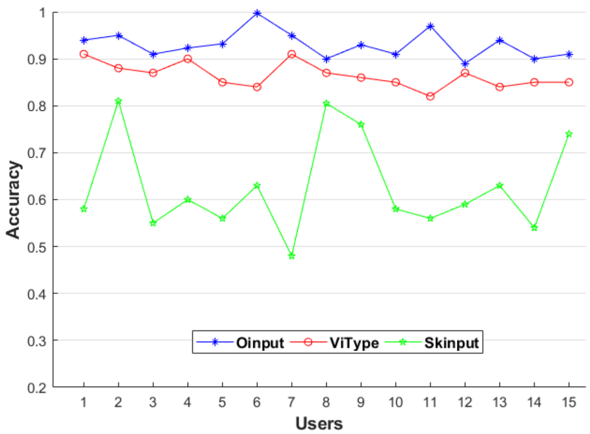


Fig. 9. Fifteen volunteers asked to tap the specified 100 buttons to calculate the average accuracy.

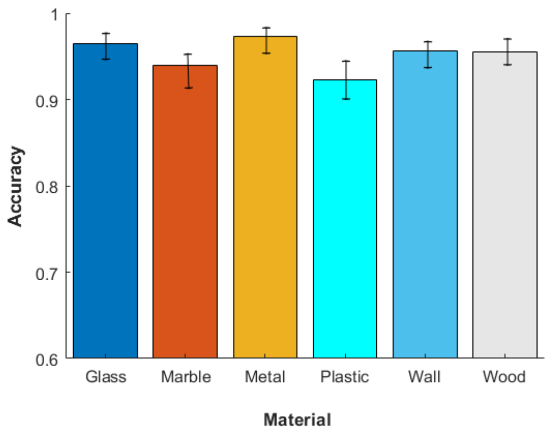


Fig. 10. The average precision achieved on the plane of different materials hitting the specified 100 buttons on the trained system.

number as the training set, and compare the accuracy of inputting the specified letters to verify the superiority of the system.

- We asked 30 volunteers to use Oinput on a desktop with a paper keyboard. Volunteers tapped each of the 31 buttons on the paper keyboard 100 times according to their usual input habits. The system randomly takes 30 samples for each button as a training set. Surprisingly, as shown in Figure 7, the average recognition rate of the test set exceeds 99%, and the highest recognition rate reaches 100%.
- We take the same training method as the first step for other systems. Each button randomly takes 30 samples as a training set. After comparison, as shown in Figure 8, we are more prominent in the practicality of the input. We believe that the main reason is that ViType is difficult to operate, even if the position is marked on the back of the hand, there will still be a slip when tapping. Also, ViType needs to be tapped three times when entering ‘C’, which gives a greater chance of misjudgment.
- Because it takes Oinput to train 31 buttons, ViType only needs to train 9 in the case of marking. Therefore, under the premise of the same user experience, we need to reduce the number of training sets of Oinput to 1/3, that is, 10 training sets. Compare the recognition accuracy of our system with other systems. We found that, as shown in

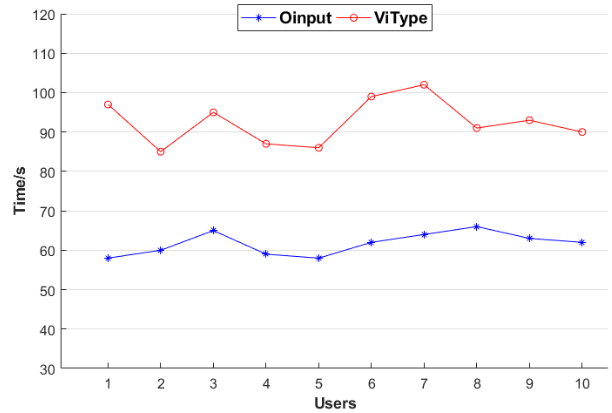


Fig. 11. The time spent by volunteers in completing the training sample collection in the initial scene of the system.

Figure 9, our accuracy is still much higher than the two systems, the highest is 100%, and the average recognition accuracy is 93.3%. The comparison shows that Oinput is superior in performance.

C. *Convenience of the Oinput system*

In the convenience experiment, we assume that the system is in the initial situation, such as the user just bought, need to enter the training set into the system. If the system needs to enter the training sample for a particularly long time, then the user will feel very troublesome and the product experience will be worse. We try to measure the time it takes to complete the entry of the training set, according to the instructions of the two systems.

- Oinput requires the user to tap 10 times for each button. However, the ViType system requires the user to train at four positions (up, down, left, and right) around the center point of the nine buttons, and that it is also required to tap at each different force. You need to enter 30 samples each time. For convenience, we allow volunteers to strictly control the intensity of their own taps, and only 10 samples per point. As shown in Figure 11, Oinput hits 310 times, which takes about 60 seconds. ViType needs to tap 450 times, which takes about 85-100 seconds, which is about 1.5 times that of Oinput.

D. *Robustness of the Oinput system*

The last step is to verify the robustness of the system, such as whether it will be affected in a noisy environment, whether it can work properly on different planes, and how much training set is required for the general input to be completed.

(1) We found that the bar is a very good experimental venue. The bar is very noisy, this is exactly what we want. At the same time, the bar has a glass table, a marble bar, a metal end plate, a plastic table, wooden finishes and limestone walls. Therefore, we are testing in a noisy environment whether different planes will affect the accuracy of the recognition. As shown in Figure 10, we tapped the specified button 100 times, and the recognition accuracy of different materials is greater than 90%.

(2) We know from the previous experiment, as shown in Figure 12, if ViType is to achieve accurate input without marking, it requires a large training set as a support. So it can be seen that when Oinput utilizes less than 20 training sets, it can already

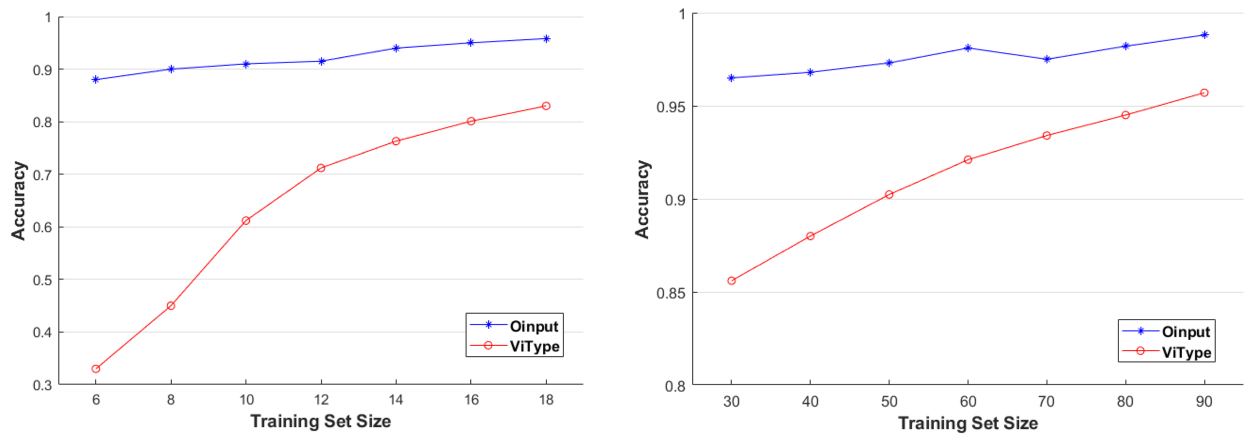


Fig. 12. 30 volunteers in the absence of the identification, tap the specified 9 buttons (A, D, G, J, M, P, T and W) 300 times, calculate the average precision tha can be achieved by randomly extracting the number of training samples.

achieve more than 90% accuracy.

VI. Conclusion

Oinput is a novel text input system that attempts to discard physical keyboards. Oinput uses a small, inexpensive vibration sensor that can be embedded in a watch and a low-energy method to achieve high-precision recognition. Regardless of whether there is a paper keyboard on the desktop, the user only needs to tap the layout of the QWERTY keyboard, and Oinput can recognize the user's input. Oinput maintains user-friendliness while achieving strong robustness. Regardless of whether the user is at different strengths or in different locations, Oinput can accurately identify the input content for the user. What's more, Oinput's adaptive function can maintain high accuracy in the environment or the user's habits change over time.

VII. Acknowledgement

The research was supported in part by the China NSFC 61872246, 61502313, 61872248, 61472259, 61836005, 61601308 , Joint Key Project of the National Natural Science Foundation of China (Grant No. U1736207), Guangdong Natural Science Foundation 2017A030312008, Shenzhen Science and Technology Foundation JCYJ20170817095418831, JCYJ20170302140946299, JCYJ20170412110753954, Fok Ying-Tong Education Foundation for Young Teachers in the Higher Education Institutions of China(Grant No.161064), Guangdong Talent Project 2015TX01X111 and GDUPS (2015), Tencent "Rhinoceros Birds"-Scientific Research Foundation for Young Teachers of Shenzhen University.

References

[1] Nandakumar, Rajalakshmi, et al. "FingerIO:Using Active Sonar for Fine-Grained Finger Tracking." *CHI Conference on Human Factors in Computing Systems* ACM, 2016:1515-1525.

[2] Wang, Wei, A. X. Liu, and K. Sun. "Device-free gesture tracking using acoustic signals." *International Conference on Mobile Computing and NETWORKING* ACM, 2016:82-94.

[3] Gong, Jun, X. D. Yang, and P. Irani. "WristWhirl: One-handed Continuous Smartwatch Input using Wrist Gestures." *Symposium on User Interface Software and Technology* ACM, 2016:861-872.

[4] Gordon, Mitchell, T. Ouyang, and S. Zhai. "WatchWriter:Tap and Gesture Typing on a Smartwatch Miniature Keyboard with Statistical Decoding." *CHI Conference on Human Factors in Computing Systems* ACM, 2016:3817-3821.

[5] Wang, Junjue, et al. "Ubiquitous keyboard for small mobile devices:harnessing multipath fading for fine-grained keystroke localization." *International Conference on Mobile Systems, Applications, and Services* ACM, 2014:14-27.

[6] Liu, Jian, et al. "Snooping Keystrokes with mm-level Audio Ranging on a Single Phone." (2015):142-154.

[7] Yin, Yafeng, et al. "CamK: Camera-Based Keystroke Detection and Localization for Small Mobile Devices." *IEEE Transactions on Mobile Computing* 17.10(2018):2236-2251.

[8] L. Wang, H. Yang, J. Long, K. Wu, J. Chen. "Enabling ultra-dense uav-aided network with overlapped spectrum sharing: Potential and approaches." *IEEE Network* 32 (5) (2018) 85–91.

[9] Harrison, Chris, D. Tan, and M. Dan. "Skinput:appropriating the body as an input surface." *Sigchi Conference on Human Factors in Computing Systems* ACM, 2010:453-462.

[10] Chen, Wenqiang, et al. "ViType: A Cost Efficient On-Body Typing System through Vibration." *IEEE International Conference on Sensing, Communication, and NETWORKING* IEEE, 2018.

[11] Shahriar Nirjon, Jeremy Gummeson, Dan Gelb, and Kyu-Han Kim. 2015. TypingRing: A Wearable Ring Platform for Text Input. In Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '15). ACM, New York, NY, USA, 227-239. DOI: https://doi.org/10.1145/2742647.2742665

[12] Çağdaş Karataş and Marco Gruteser. 2015. Printing multi-key touch interfaces. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15). ACM, New York, NY, USA, 169-179. DOI: https://doi.org/10.1145/2750858.2804285

[13] Abdelnasser, Heba, M. Youssef, and K. A. Harras. "MagBoard: Magnetic-Based Ubiquitous Homomorphic Off-the-Shelf Keyboard." *IEEE International Conference on Sensing, Communication, and NETWORKING* IEEE, 2016.

[14] Zhenjiang Li, et al. "iType: Using eye gaze to enhance typing privacy." *INFOCOM 2017 - IEEE Conference on Computer Communications, IEEE* IEEE, 2017:1-9.

[15] Vidya Lakshmipathy, Chris Schmandt, and Natalia Marmasse. 2003. TalkBack: a conversational answering machine. In Proceedings of the 16th annual ACM symposium on User interface software and technology (UIST '03). ACM, New York, NY, USA, 41-50. DOI: https://doi.org/10.1145/964696.964701

[16] Singh, V. R., S. Yadav, and V. P. Adya. "Role of natural frequency of bone as a guide for detection of bone fracture healing." *Journal of Biomedical Engineering* 11.6(1989):457-461.

[17] Singh, V. R., et al. "A stress wave propagation technique for bone repair study. " *IEEE transactions on bio-medical engineering* 37.10(1990):1014.

[18] Pelker, R. R., and S. Saha. "Stress wave propagation in bone. " *Journal of Biomechanics*16.7(1983):481-489.

[19] Saha, Subrata, and R. S. Lakes. "A non-invasive technique for detecting stress waves in bone using the piezoelectric effect. " *IEEE transactions on bio-medical engineering* 24.6(1977):508.

[20] Yu-Yong, H. U., L. V. Ming, and S. Y. Wang. "Measurement of Material Natural Frequency and Elastic Modulus by Knocking Method." *Mechanical Engineering & Automation* (2010).