深圳大学
SHENZHEN UNIVERSITY

深圳大学物联网研究中心
The IoT Research Center

FaceInput: A Hand-Free and Secure Text Entry System through Facial Vibration

Presenter: Yandao Huang

Maoning Guan, Wenqiang Chen, Yandao Huang,
Rukhsana Ruby, Kaishun Wu

Shenzhen University, Shenzhen, China

SECON' 19, Boston

2019/6/12

Background

The screen size is getting smaller, and so it is the input interface which makes the interactive experience poorer.

Background

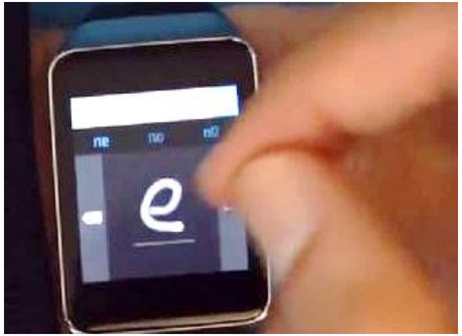
There are many applications in the wearable devices, which require the text input interface.

Existing Input Methods

01 Simple Typing

Small Size
User Unfriendly

02 Finger Tracking

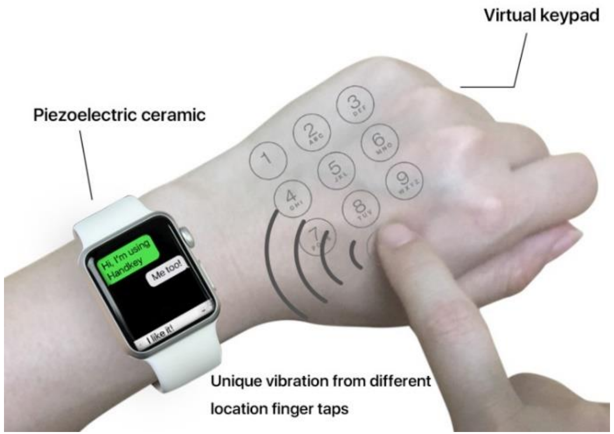


Too Slow

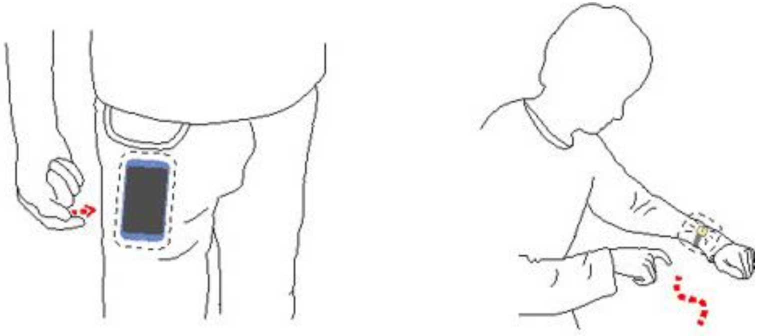
03 Speech Input

Disturbing
Poor Noise Resistance
Sensitive Information

Existing Input Methods



ViType[1]

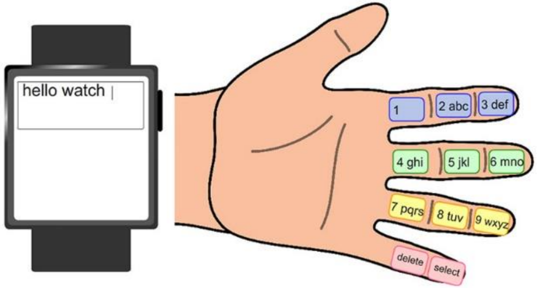


FingerIO[2]

Opposite-side interaction is not available for users whose hands are fully occupied with other tasks.

1. Wenqiang Chen, et.al. ViType: A Cost Efficient On-Body Typing System through Vibration, IEEE SECON, 2018.
2. Rajalakshmi Nandakumar, et.al. FingerIO: Using active sonar for fine-grained finger tracking, ACM CHI, 2016.

Existing Input Methods



Float [1]

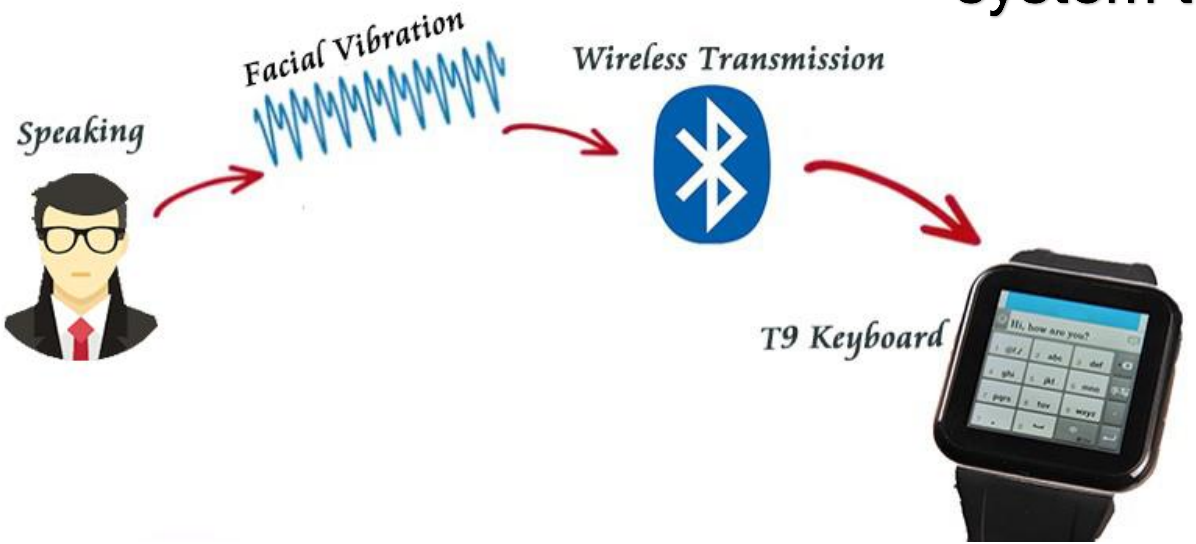
FingerT9 [2]

One-handed interaction is also not available for users whose hands are fully occupied with other tasks.

1. Ke Sun, et.al. Float: One-Handed and Touch-Free Target Selection on Smartwatches, ACM CHI, 2017.
2. Pui Chung Wong, et.al. FingerT9: Leveraging thumb-to-finger interaction for same-side-hand text entry on smartwatches, ACM CHI, 2018.

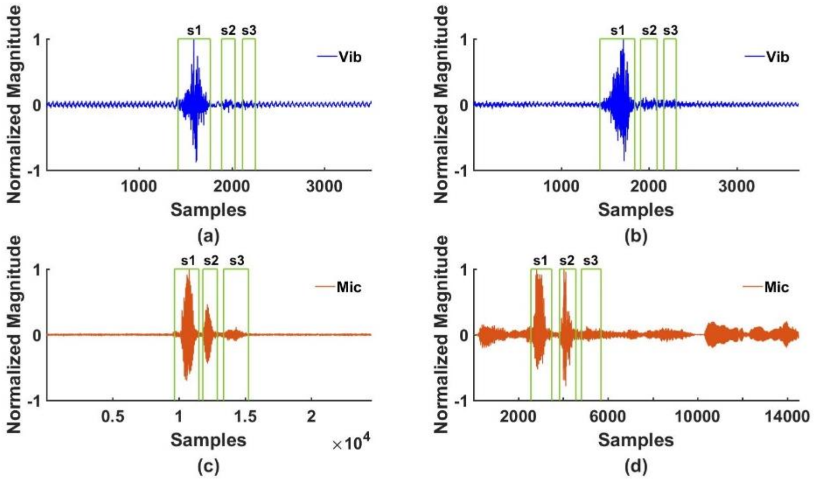
How to overcome the limitations of a small screen for smart watches with a hand-free interaction?

Text entry system——FaceInput

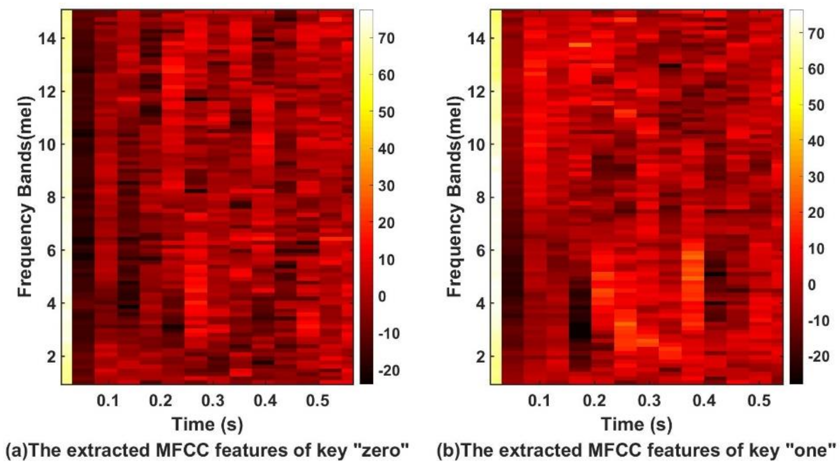


Hand-free and secure text input system through **facial vibration.** (T9 layout)

Observation



Sample raw input signals of a word(i.e., “taps”).



Example of the extracted MFCC features.

Speaking different keys (e.g., 0,1,2,…,9) ➡unique vibration profile

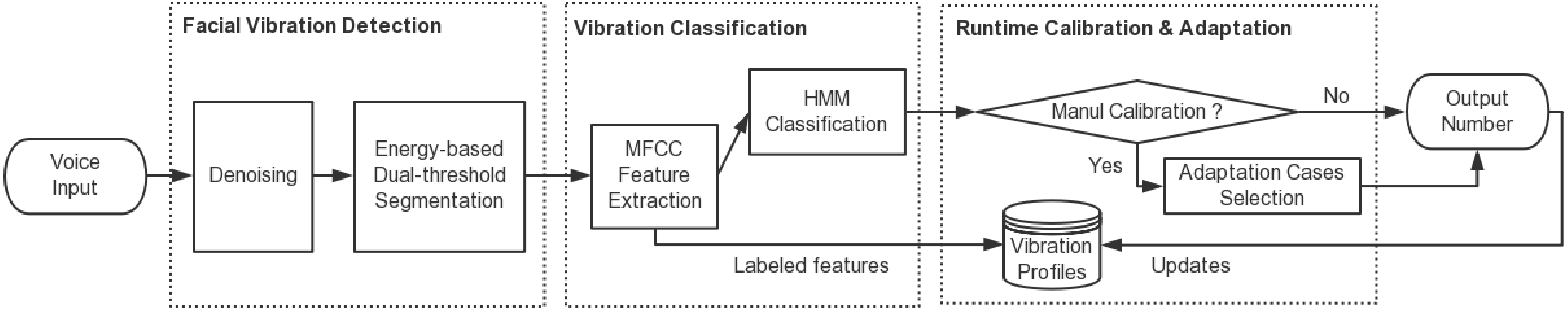
Design Goals and Challenges

1. A available system in most of the daily user scenarios.

2. Robust enough to give the correct output when some variations occur.

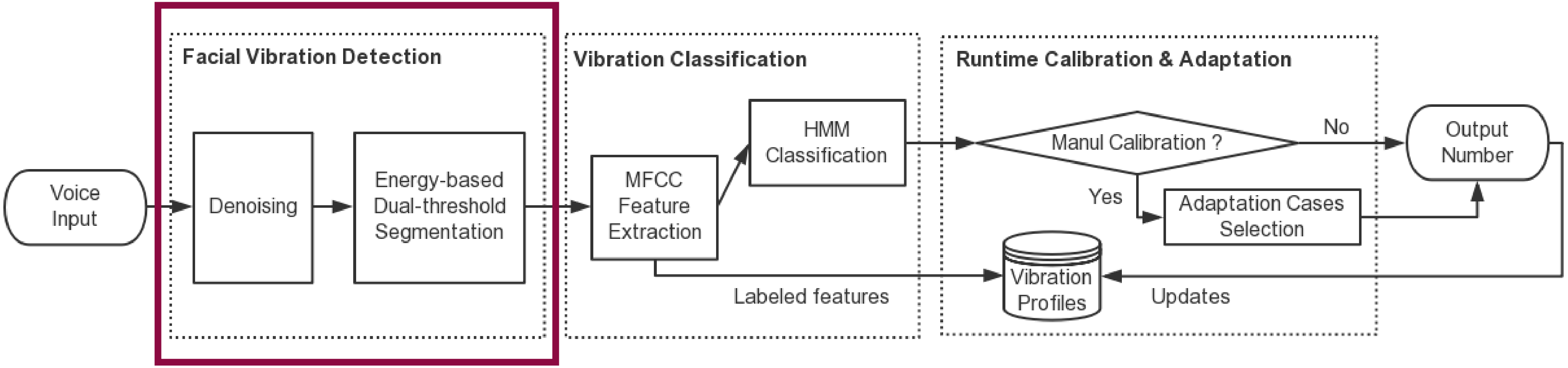
3. Efficient with low time and computation overhead.

Architecture



The architecture of FaceInput.

Architecture

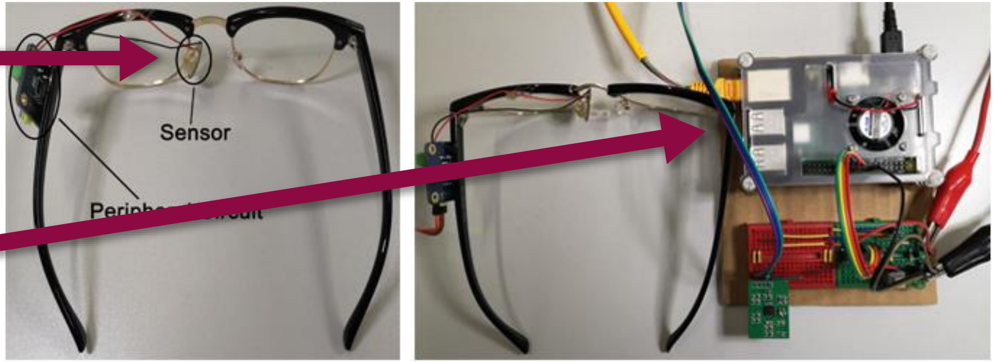


The architecture of FaceInput.

Sensing | **Denoising** | **Segmentation**

1. Piezoelectric ceramic sensor
   Diameter: 20 mm
   Thickness: 0.4 mm

2. A Raspberry Pi with an ADC



(a) (b)

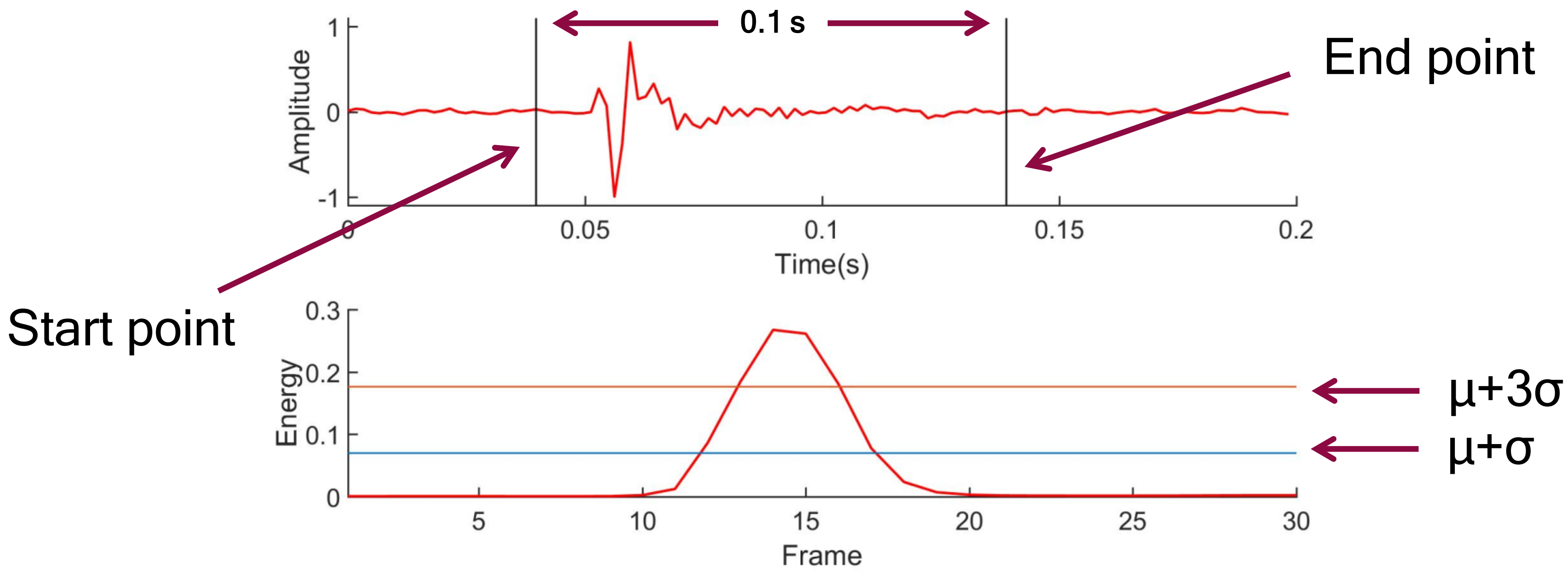
Sensing | **Denoising** | Segmentation

Human mobility(e.g., walking)

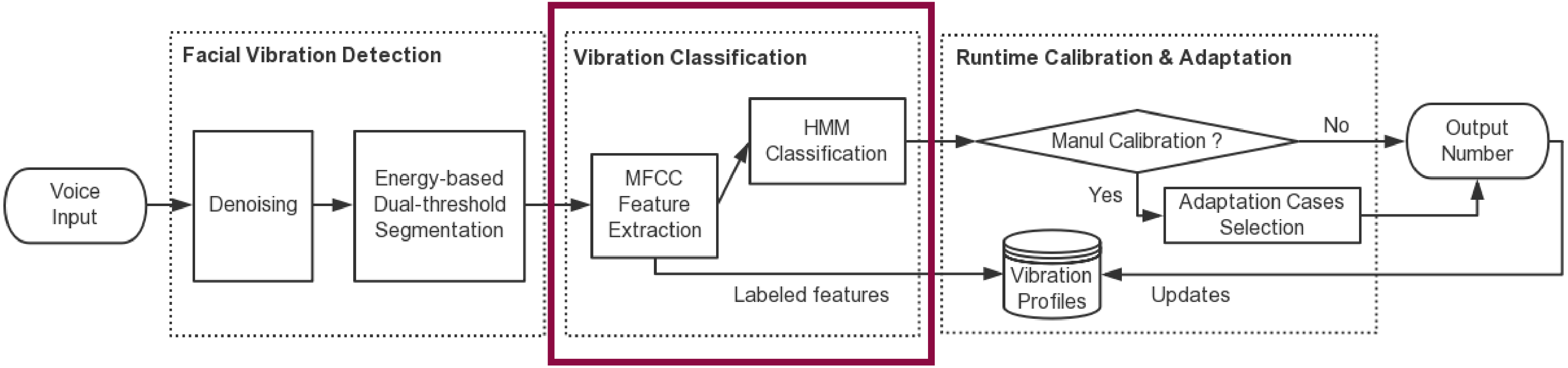
Human mobility(e.g., walking)

A Butterworth band pass filter in the 10 to 1000 Hz range.

- To Remove the low-frequency noise caused by DC & human mobility(less than 10Hz) and high-frequency noise.

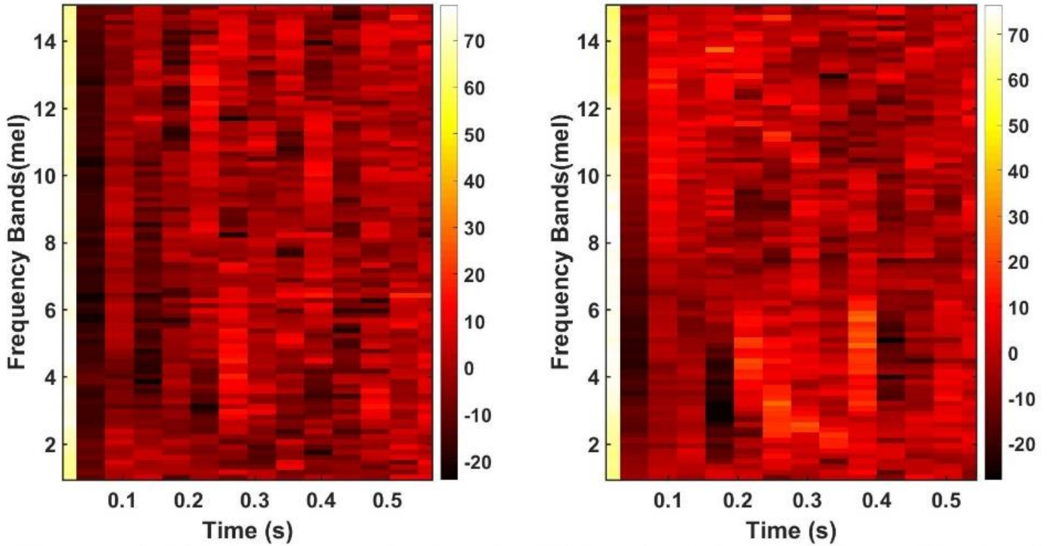

0.1 s
End point
Start point
Amplitude
1
0
-1
0
0.05
0.1
0.15
0.2
Time(s)
Energy
0.3
0.2
0.1
0
5
10
15
20
25
30
Frame
μ+3σ
μ+σ

Architecture



The architecture of FaceInput.

Feature Extraction——MFCC



(a)The extracted MFCC features of key "zero" (b)The extracted MFCC features of key "one"

Example of the extracted MFCC features.

Information in the **Time Domain** + Information in the **Frequency Domain** = **MFCC** features

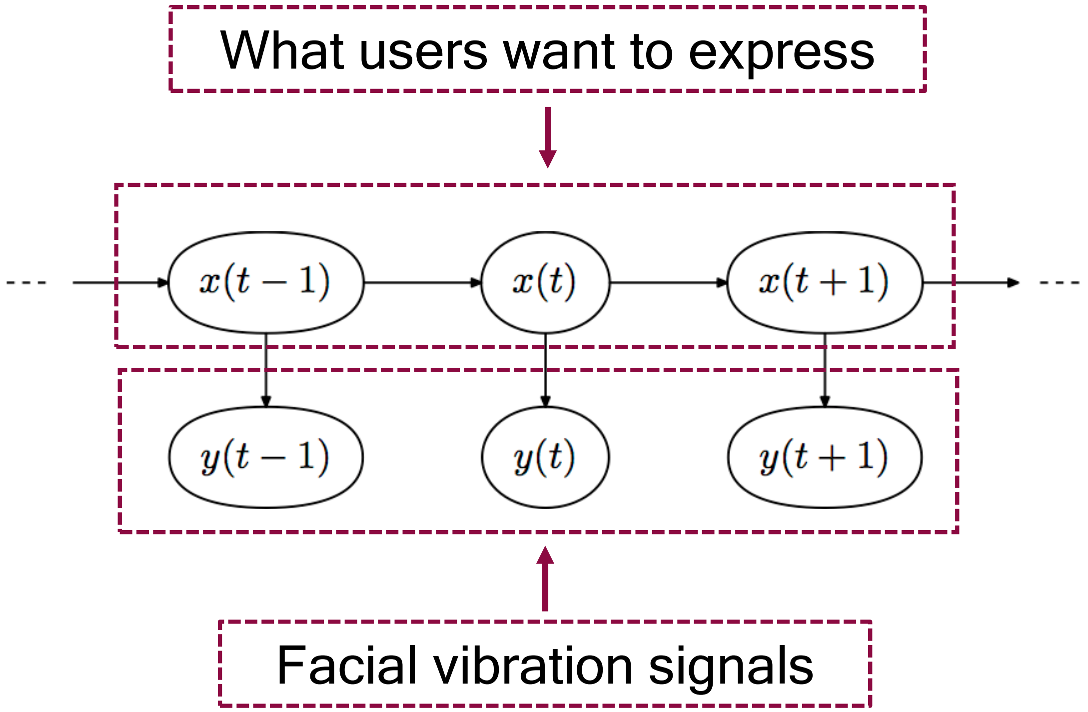
Classification Algorithm

Hidden Markov Model (HMM)

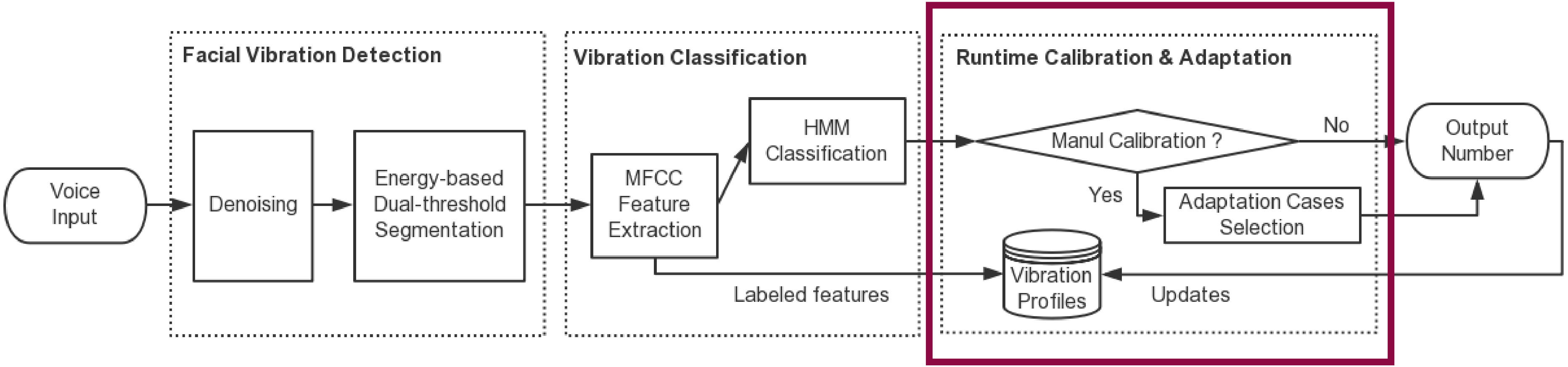
The number of discrete hidden states is set to 3 in our case.

hidden-node sequences

Observable time-varying sequences

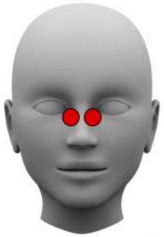


Architecture



The architecture of FaceInput.

Runtime Calibration and Adaptation

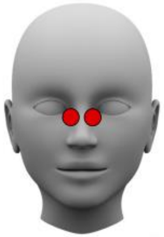


Sensor Displacement
on the face

Voice strength variation

How to deal with the deviation?

Runtime Calibration and Adaptation



Sensor Displacement
on the face

Voice strength variation

Update with candidate keys

158		
7	3	5

Evaluation

1

Accuracy

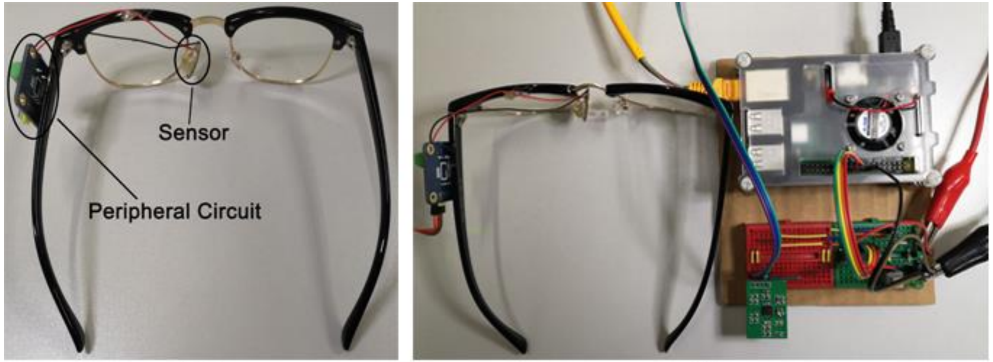
2

Robustness

3

User Study

Experimental Setup



(a) (b)

- 10 virtual keys on T9 layout
- Each participant spoke each key for 20 times
- 30 participants collected 6,000 keystrokes

Evaluation

1

Accuracy

2

Robustness

3

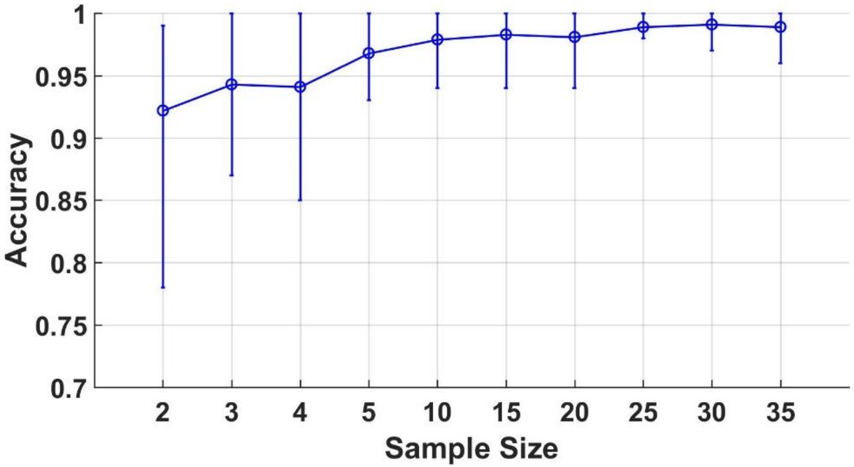
User Study

Accuracy——Baseline detection and classification

	Key0	Key1	Key2	Key3	Key4	Key5	Key6	Key7	Key8	Key9
Key0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Key1	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Key2	0.00	0.00	0.98	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Key3	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00
Key4	0.01	0.00	0.01	0.00	0.98	0.00	0.00	0.00	0.00	0.00
Key5	0.00	0.02	0.00	0.00	0.00	0.97	0.00	0.00	0.00	0.00
Key6	0.01	0.00	0.00	0.01	0.00	0.00	0.95	0.01	0.02	0.00
Key7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00
Key8	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.97	0.00
Key9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99

The average classification accuracy is 98.2%.

Accuracy——Impact of Training Set Size



Baseline accuracy: 98.2%
(training set size: 10)

Training set size enlarge from 2 to 10, the accuracy rises from 92.2% to 98.2%.

Evaluation

1

Accuracy

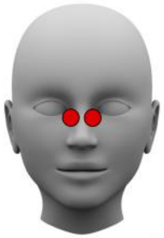
2

Robustness

3

User Study

Robustness——Variation



Sensor Displacement on the face

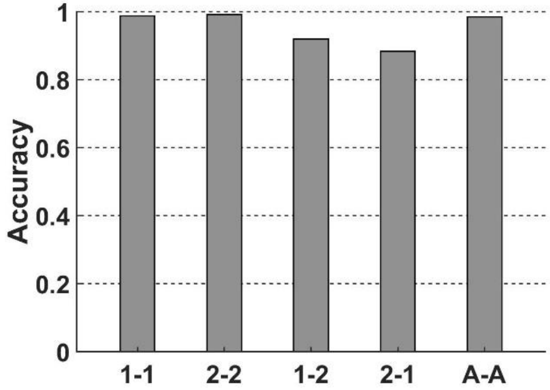


Fig. 1. Accuracy of positional variation of glasses.

Voice strength variation

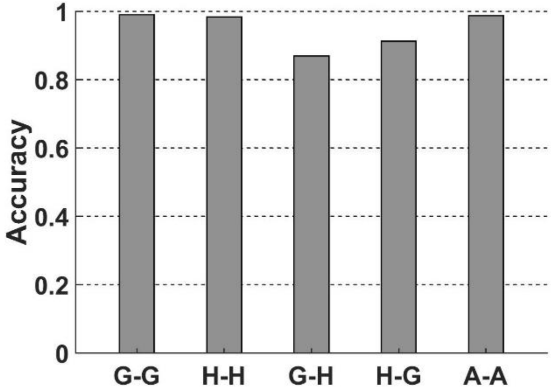


Fig. 2. Impact of different voice strength.

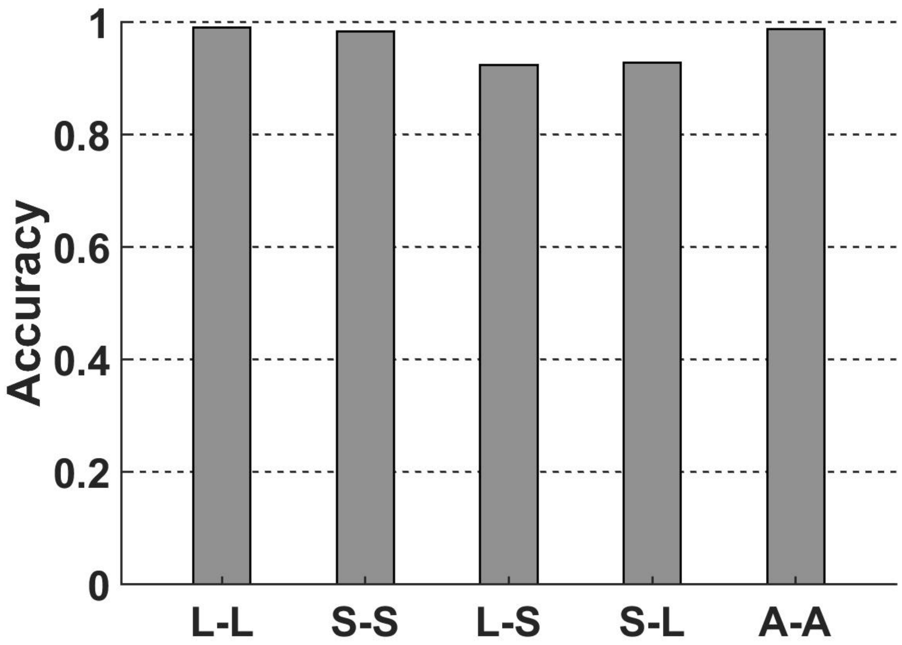
Gentle - Hard - All

X-axis format:
"training data" – " test data"

The system's performance suffers a degradation under variation impact.

Runtime Adaptation

Robustness——Voice Length

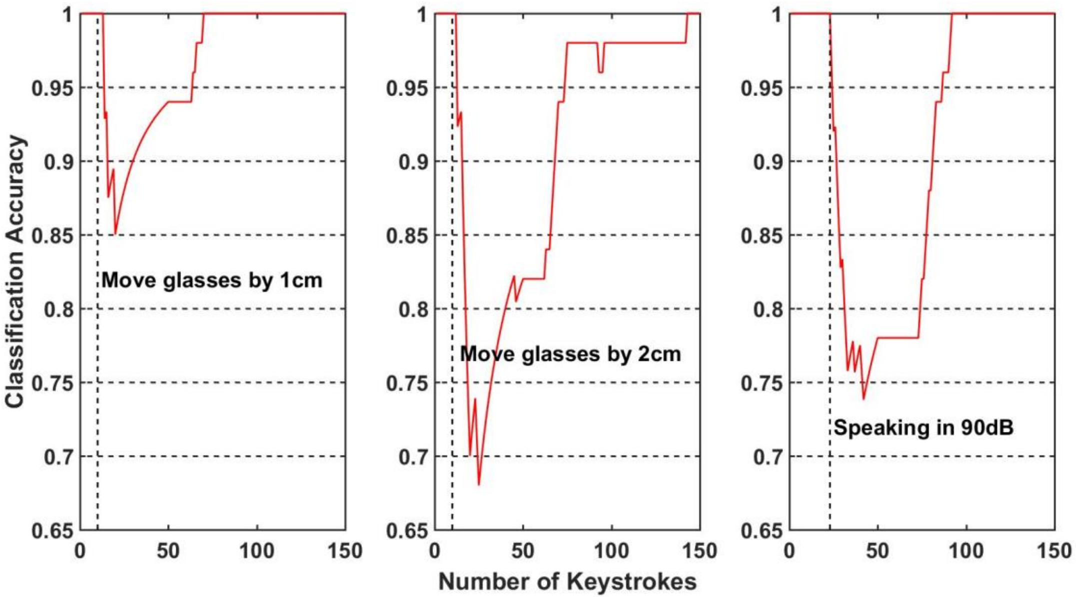


Impact of different voice length

Long - Short - All

- Different voice length hardly affect classification accuracy (above 92%), which should owe to MFCC features and HMM algorithm.
- The accuracy recovers to “All-All” situation.

Robustness——Calibration & Adaptation



(Dotted lines denote the occurrence moment of variation)

The calibration and adaptation scheme can mitigate the variation impact, and it can recover the accuracy to 100% after a few tens of inputs

Robustness——Mobility

Items	Standing (Baseline)	Walking	Shaking the head
Accuracy	98.6%	94.9%	97.1%

While walking and shaking the head, the average accuracy is **96%**, which shows the robustness to human mobility.

The noise caused by human mobility is at low frequency（less than 10Hz）, and we remove it via a Butterworth band pass filter in the 10 to 1000Hz range.

Evaluation

1

Accuracy

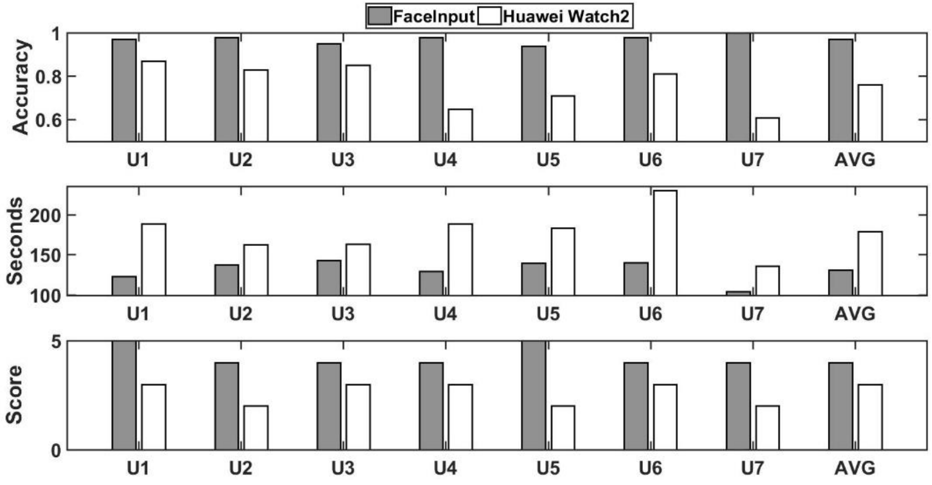
2

Robustness

3

User Study

User Study

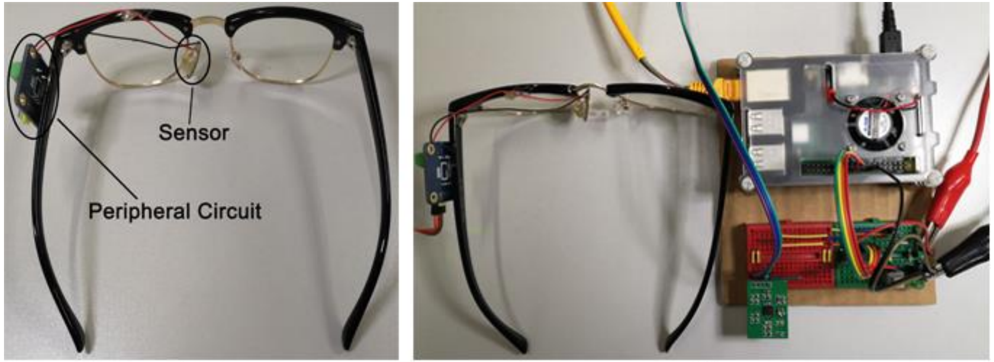


Comparison of input accuracy, speed and user experience between FaceInput and Huawei Watch2. (input 100 random numbers from 0-9.)

Items	FaceInput	HUAWEI Watch2
Accuracy	97%	76%
Input Speed(s)	131	179
Score(0-5)	4	3

FaceInput, as a text entry system with a hand-free interaction,
does provide higher input accuracy, faster input speed, and better user experience.

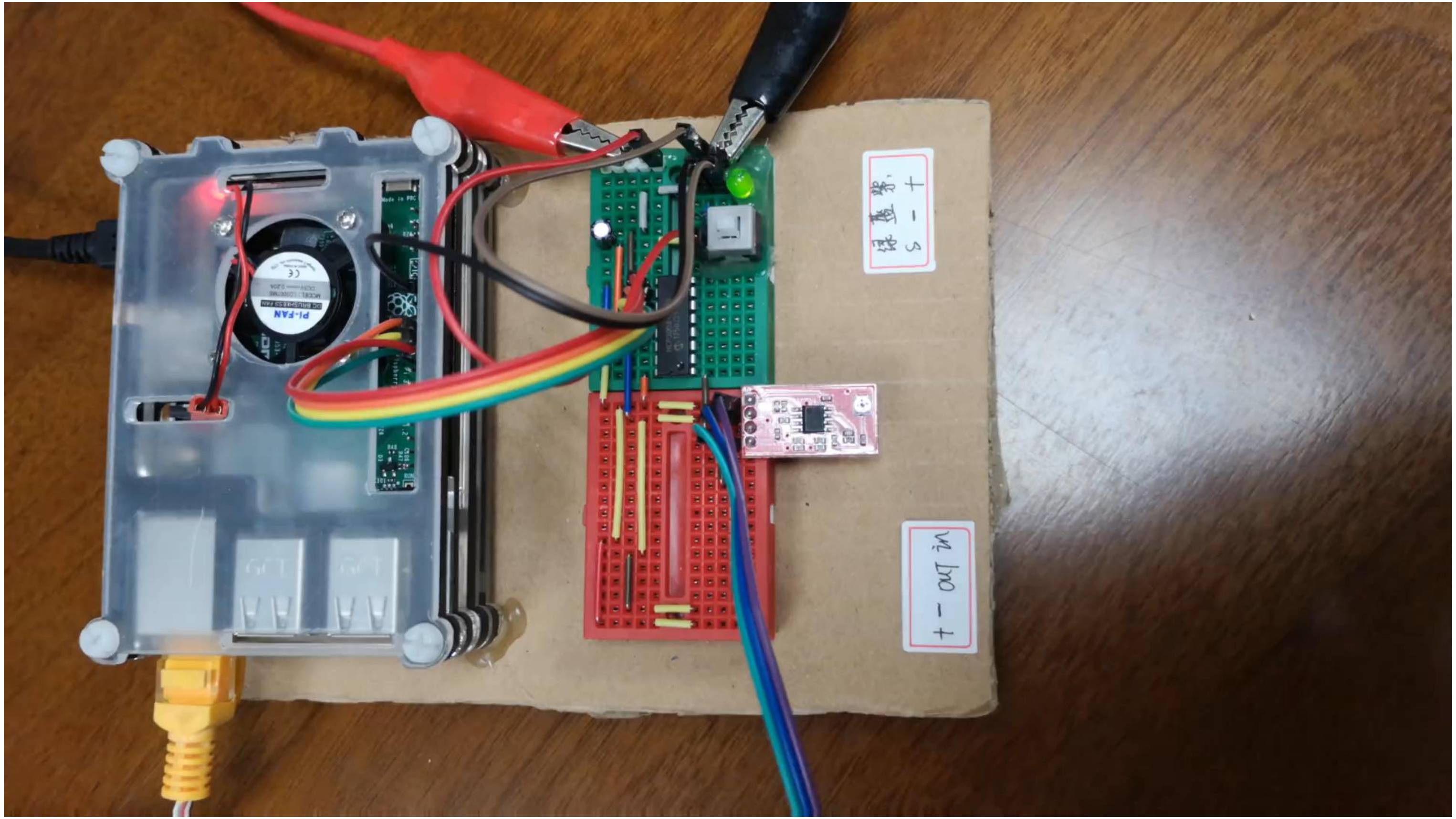
Cost



(a) (b)

- Initial training: 3 minutes
- Training duration: 2.2 seconds
- Latency : 0.25 seconds
- Sensor: 0.15 dollars.

Demonstration of FaceInput



Conclusion

- FaceInput firstly conducts a hand-free and secure text entry system via facial vibration with only one vibration sensor.
- Our system achieves high recognition accuracy for ten keys with accuracy of 98.2%.
- We evaluate the accuracy and robustness under different common conditions and design the calibration scheme to improve the robustness.
- FaceInput outperforms the input method in COTS smart watch.

Thank you!

Maoning Guan

guanmaoning2018@email.szu.edu.cn

- Different voice length?
- Secure text input?