EchoWrite: An Acoustic-based Finger Input System Without Training

Yongpan Zou[1], **Qiang Yang**[1], Rukhsana Ruby[1], Yetong Han[1], Mo Li[2], Kaishun Wu[1]

[1] Shenzhen University, China
[2] Nanyang Technological University, Singapore

July 9, 2019 @ Dallas, Texas, USA

深圳大学
SHENZHEN UNIVERSITY

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

ICDCS 2019
The 39th IEEE International Conference on Distributed Computing Systems

Outline

Motivation

PC Era

Mobile Era

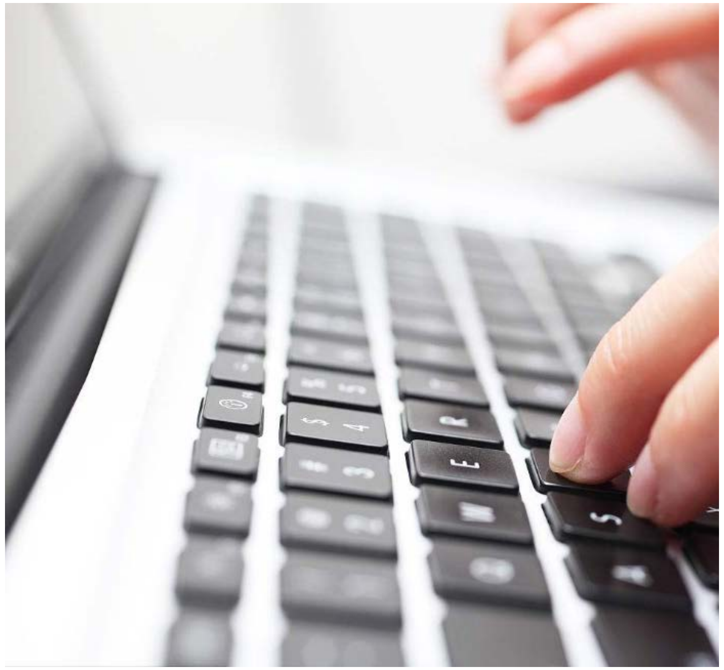
IoT Era

Motivation

Traditional interaction interface - Keyboard

Smartphone

Table computer



PC

Motivation

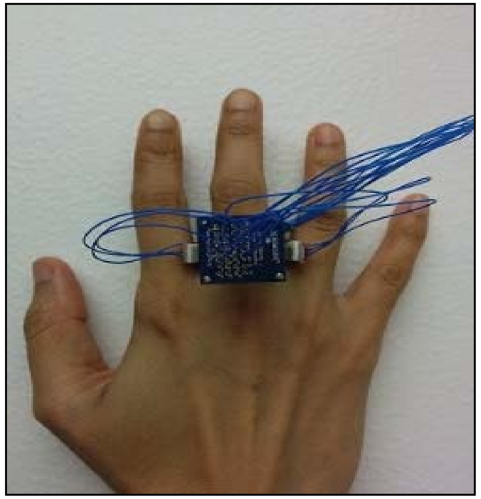
For new smart devices? **Small screen size / no screen!**

Smart watch

Smart glass

Smart home

Related work



RF

speech recognition

IMU

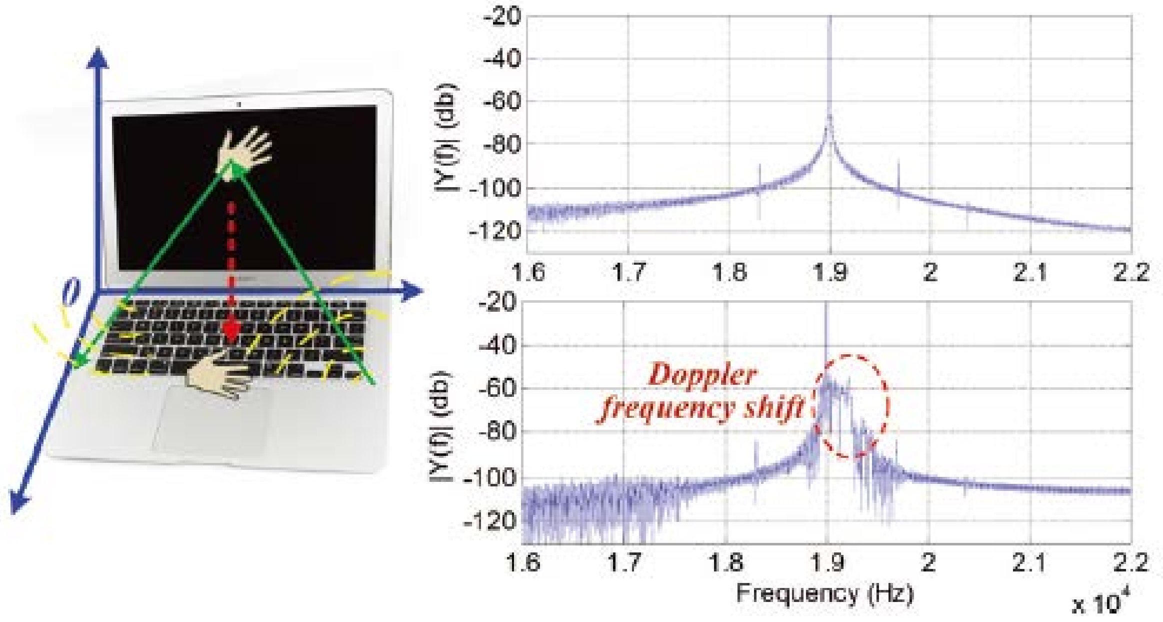
Unstable

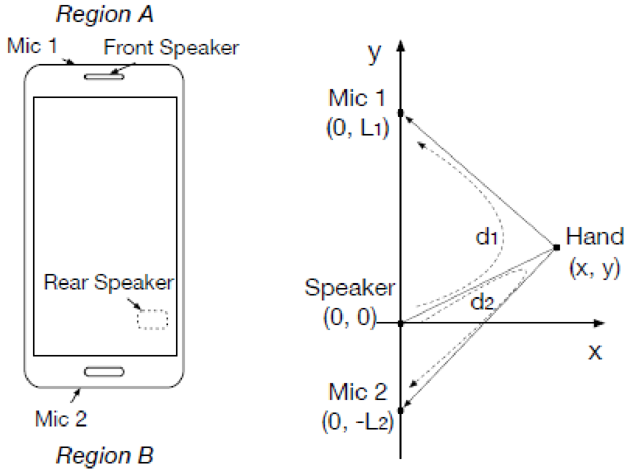
Privacy concern

Wearing device

1. L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in Proceedings of ACM MobiSys, 2015.
2. J. Wang, D. Vasisht, and D. Katabi, "RF-IDraw: virtual touch screen in the air using rf signals," in Proceedings of ACM SIGCOMM, 2014.
3. S. Nirjon, J. Gummeson, D. Gelb, and K.-H. Kim, "Typingring: A wearable ring platform for text input," in Proceedings of ACM MobiSys, 2015.
4. C. Amma, M. Georgi, and T. Schultz, "Airwriting: Hands-free mobile text input by spotting and continuous recognition of 3d-space handwriting with inertial sensors," in Proceedings of IEEE ISWC, 2012.

Related work





Hand gesture recognition

Coarse-grained HAND gesture

Acoustic finger tracking

Two microphones are required

5. *S. Gupta, D. Morris, S. Patel, and D. Tan, "Soundwave: using the Doppler effect to sense gestures," in Proceedings of ACM CHI, 2012.*
6. *W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in Proceedings of ACM Mobicom, 2016.*
7. *Y. Zou, Q. Yang, Y. Han, D. Wang, J. Cao, and K. Wu, "Acoudigits: Enabling users to input digits in the air," in IEEE PerCom, 2019*
8. *W. Ruan, Q. Z. Sheng, L. Yang, T. Gu, P. Xu, and L. Shangguan, "Audiogest: enabling fine-grained hand gesture detection by decoding echo signal," in ACM Ubicomp, 2016.*

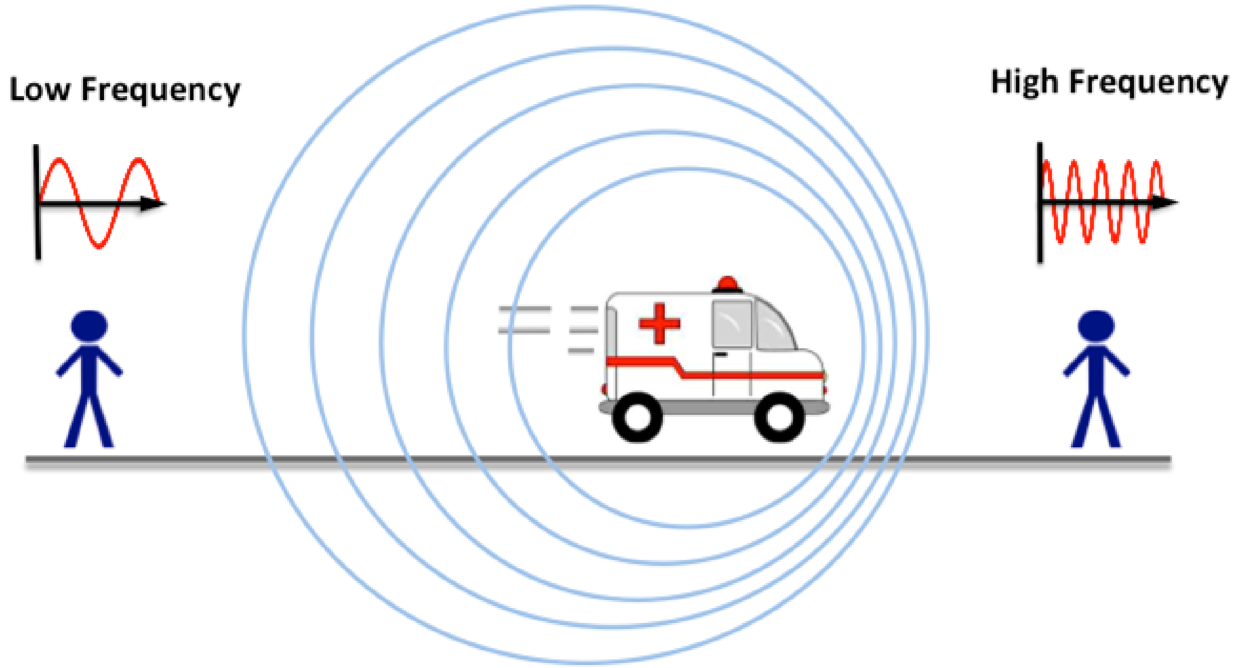
Related work

<table>
<tr><th></th><th>Device</th><th colspan="3">Device-free</th></tr>
<tr><td>Training</td><td>[3] [4]</td><td colspan="3">[7]</td></tr>
<tr><td rowspan="3">Training-free</td><td rowspan="3">[1] [2]</td><td></td><td>Coarse-gained</td><td>Fine-grained</td></tr>
<tr><td>Two mics</td><td>-</td><td>[6]</td></tr>
<tr><td>One mics</td><td>[5] [8]</td><td>EchoWrite</td></tr>
</table>

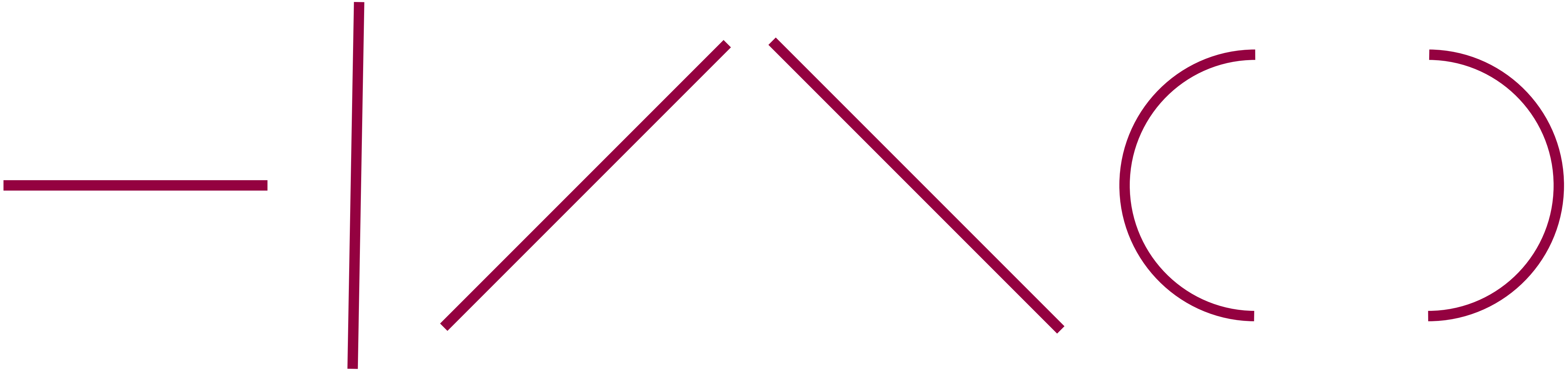
System Design - Basic idea

Principle: Doppler Effect

Frequency of a sound wave changes as a listener moves toward or away from the source



System Design - Basic idea

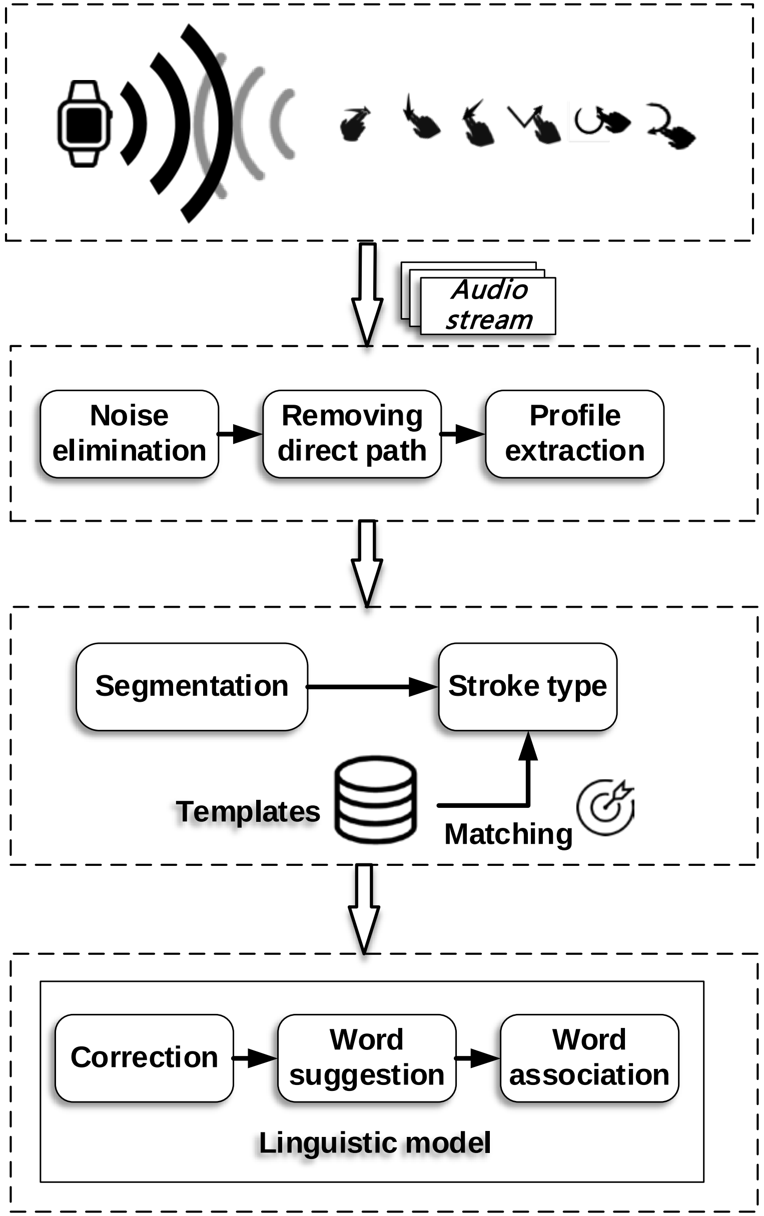


*Six basic strokes of English letters**

horizontal | vertical | left-falling | right-falling | Left-arc | Right-arc

*“*Using Mobile Phones to Write in Air*”, Sandip Agrawal , et.al, *ACM MobiSys*, 2011

System Design - framework



Data collection

Preprocessing

Stroke recognition

Text recognition

System Design - Challenges

1. *How to recognize finger gestures in a training-free way?*

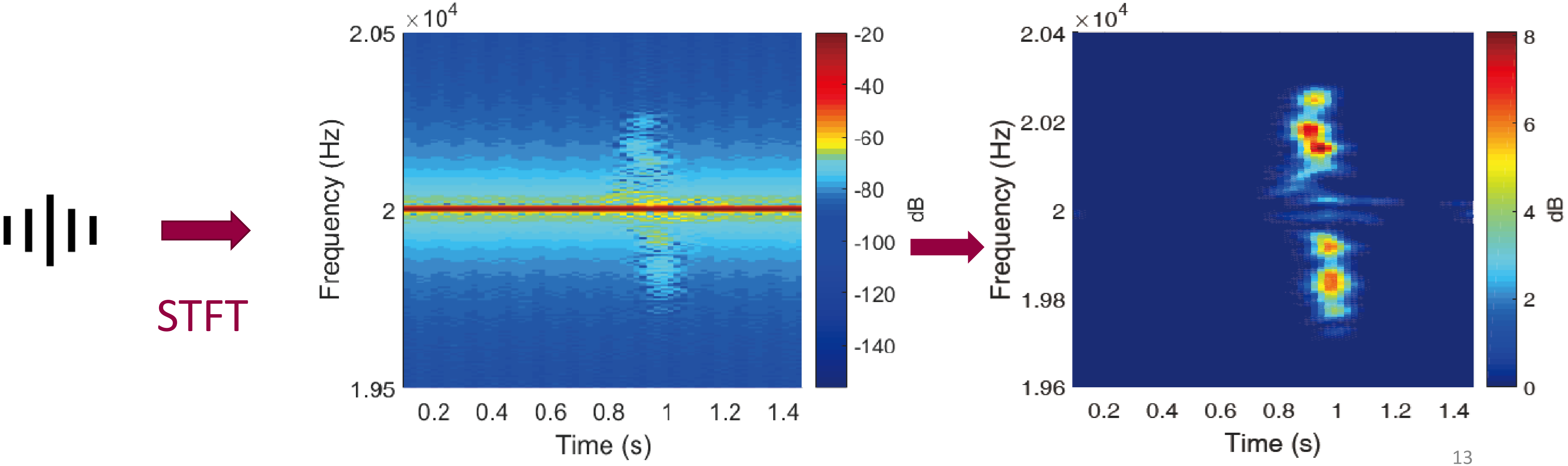
2. *How to input continuously under ambient interference ?*

3. *How to input text efficiently based on finger gestures?*

C1. How to recognize finger gestures in a training-free way?

Noise elimination

- Random noise: median filter
- Direct path: spectrum subtraction

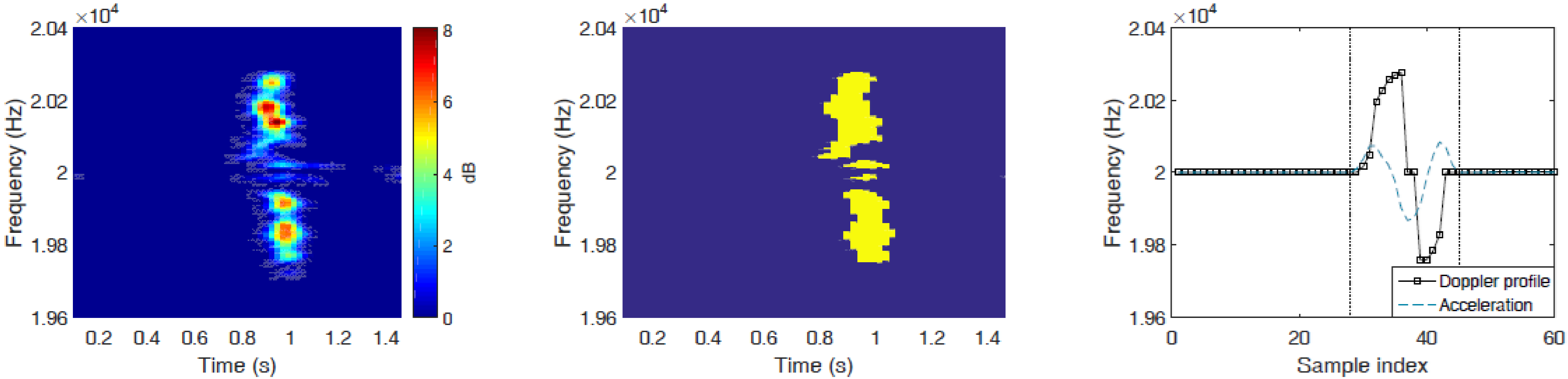


C1. How to recognize finger gestures in a training-free way?

Stroke profile extraction

- Normalization + Binarization
- Image processing: *Area open, flood fill*
- Profile extraction: Mean Value-based Contour Extraction (MVCE) algorithm

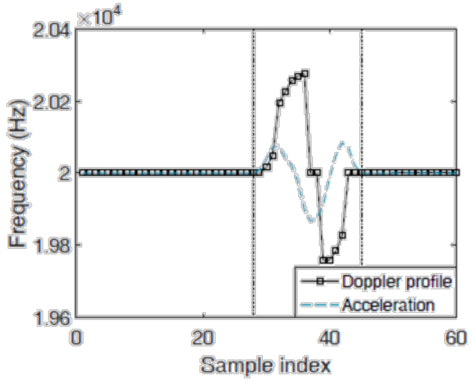
Stroke profile



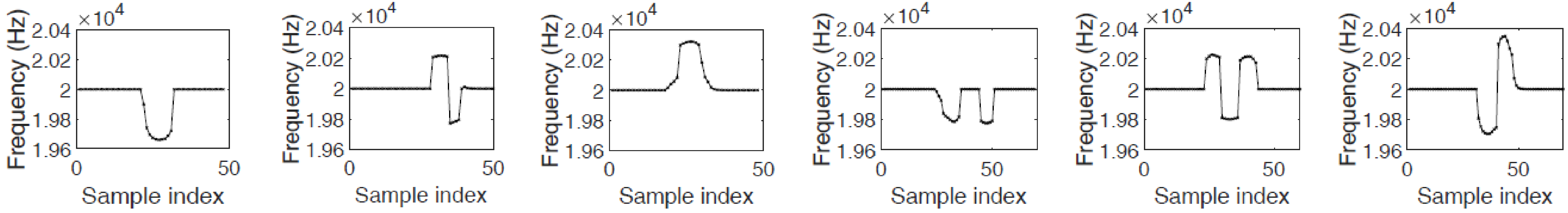
C1. How to recognize finger gestures in a training-free way?

Strokes Recognition

- Template matching
- Dynamic time wrapping (DTW)



DTW



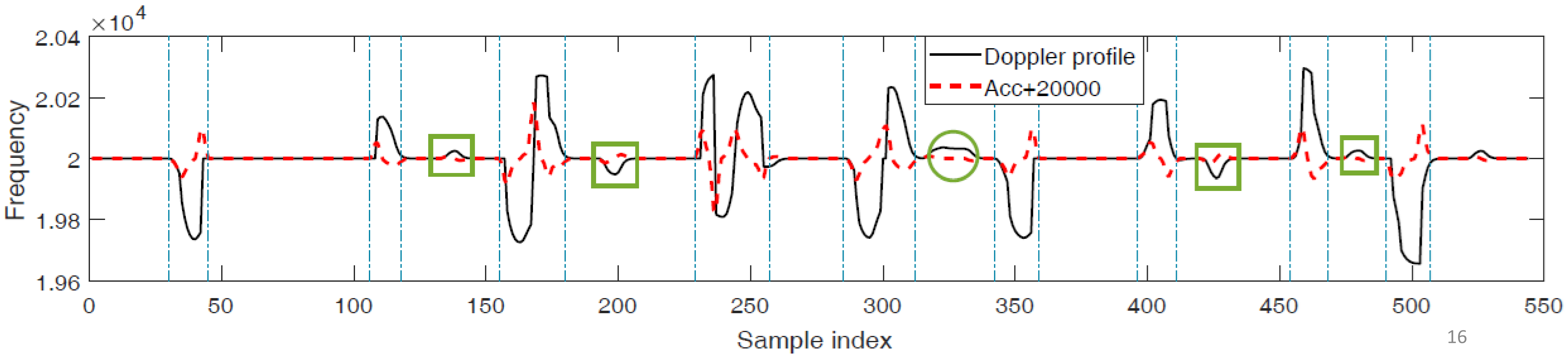
(a) S_1 (b) S_2 (c) S_3 (d) S_4 (e) S_5 (f) S_6

S_2

C2. How to input continuously under ambient interference ?

Successive Strokes Segmentation:

- Traditional methods: based on speed (i.e. Doppler frequency)
- **Key observation:** In the end of stroke writing, the speed remains but acceleration decreases notably.
- Acceleration can be utilized to discriminate strokes and other movement (arm, body, and other objects)

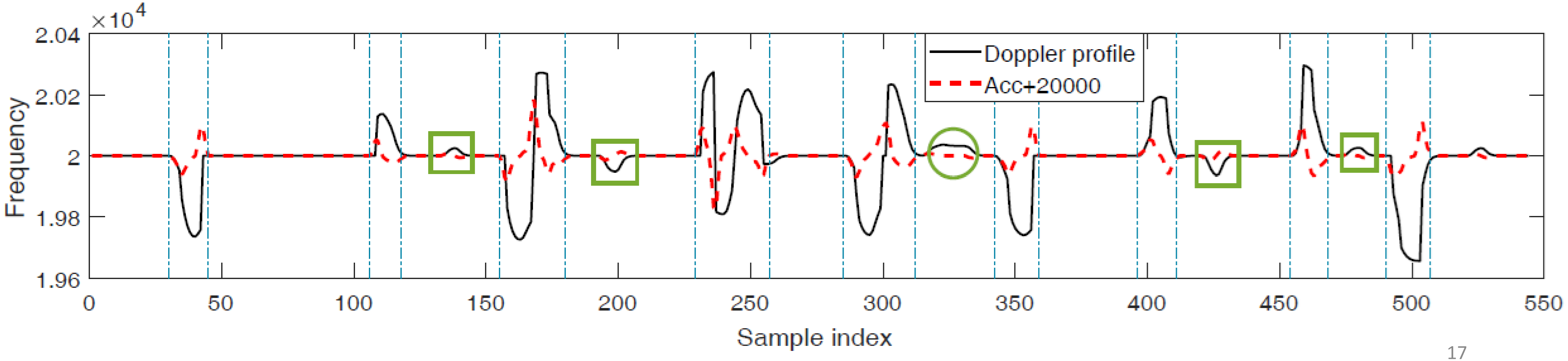


Successive Strokes Segmentation

- Acceleration: first-order difference
- Green box: moving objects
- Green circle: Uninterrupted writing

$$f_t = \frac{1 \pm \frac{v_f}{v_s}}{1 \mp \frac{v_f}{v_s}} f_0 = \pm \frac{2 f_0 v_f}{v_s \mp v_f} + f_0 \approx \pm \frac{2 f_0 v_f}{v_s} + f_0$$

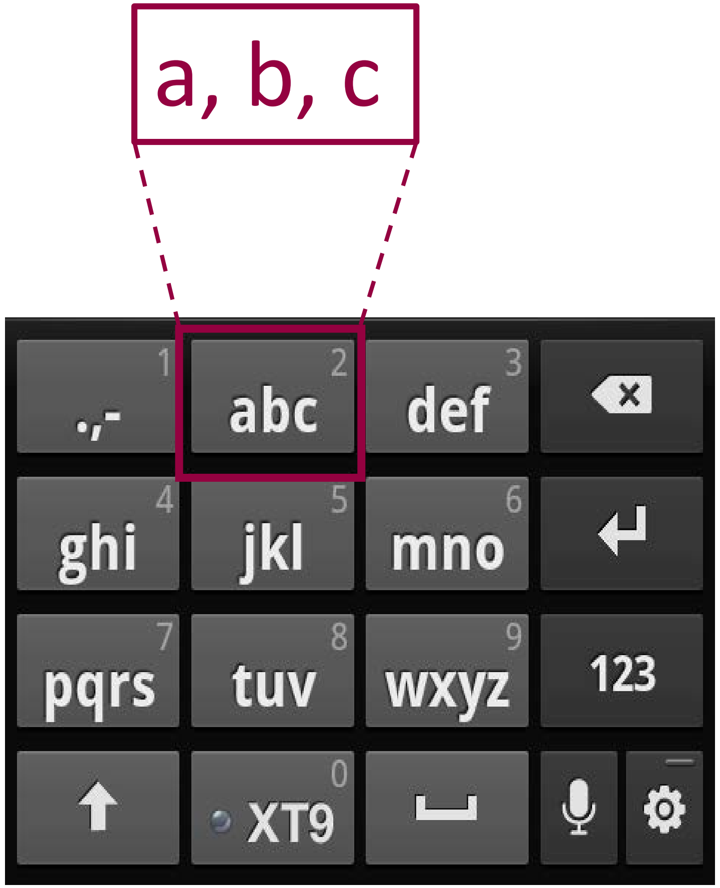
$$f_t' = \frac{2 f_0}{v_s} v_f' = \frac{2 f_0}{v_s} a$$

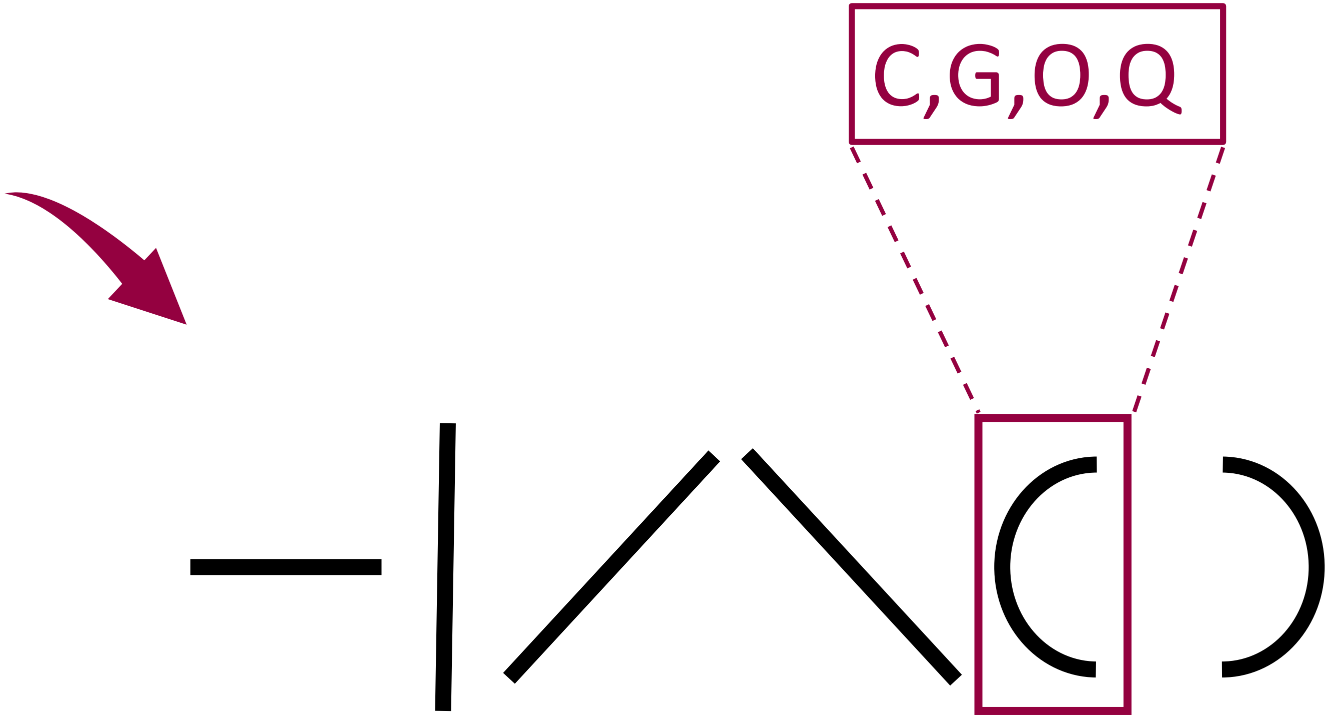


C3. How to input text efficiently based on finger gestures?

Referring to T9 Keyboard

- Each stroke represents several letters (started by this stroke when writing).
- Users can customize their own writing habits.





Example: Input 'TO'

I, T, Z, J

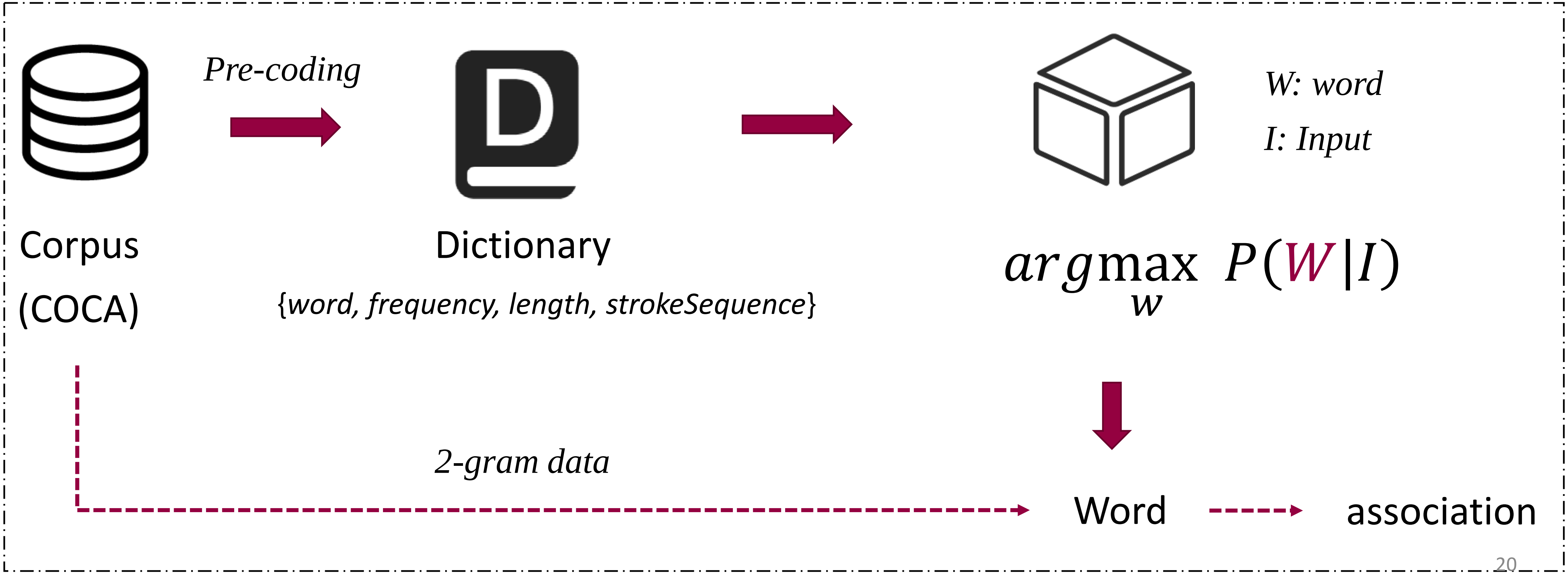
C, G, O, Q

IC
IG
IO
IQ
TC
TO ☑
TQ
…

Bayesian-based Linguistic Model

- Tolerating possible errors (recognition or writing)
- Auto-associating the next possible word.

Stroke sequence *(I)*



Evaluation - setup

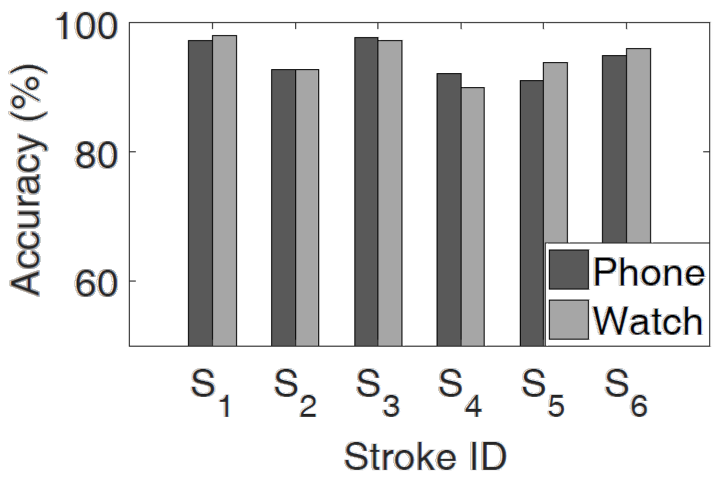
Meeting room	Lab area	Entertainment zone

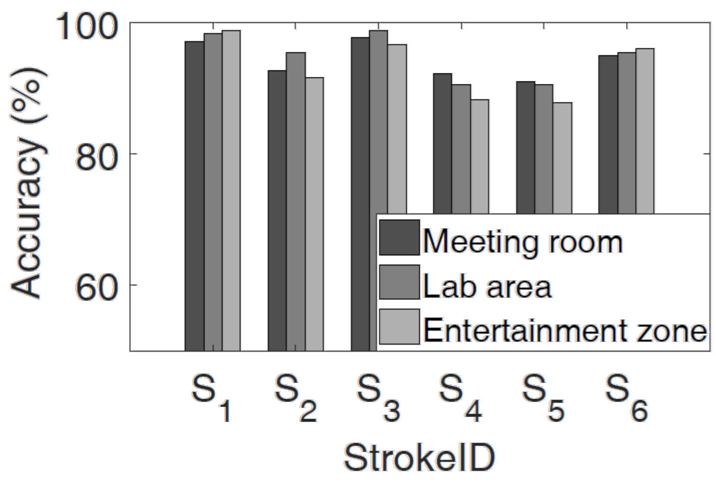
3(settings) X 6(participants)X 6(strokes)X 30(repetitions) = 3240 *strokes*

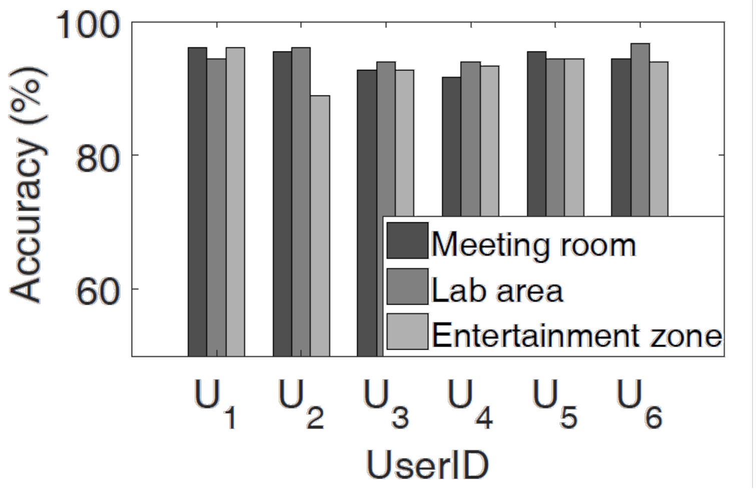
6(participants) X 10(words) X 20(repetitions) = 1200 *words*

Huawei watch 2 vs Huawei mate 9

Evaluation – stroke recognition







Different devices

- Smartwatch: 94.4%, Smartphone: 94.7%

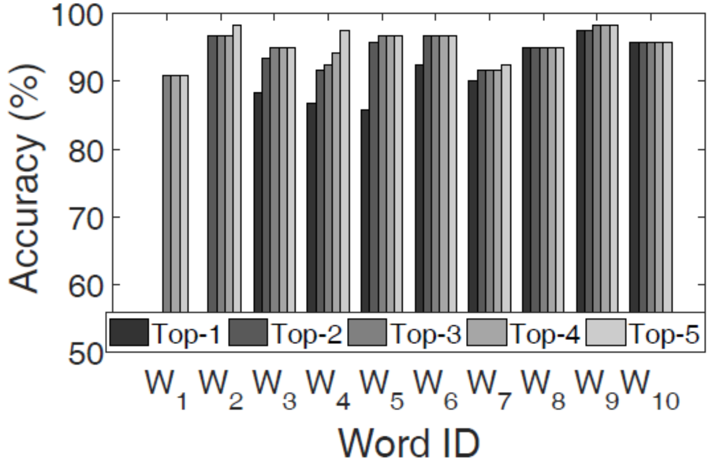
Different environment

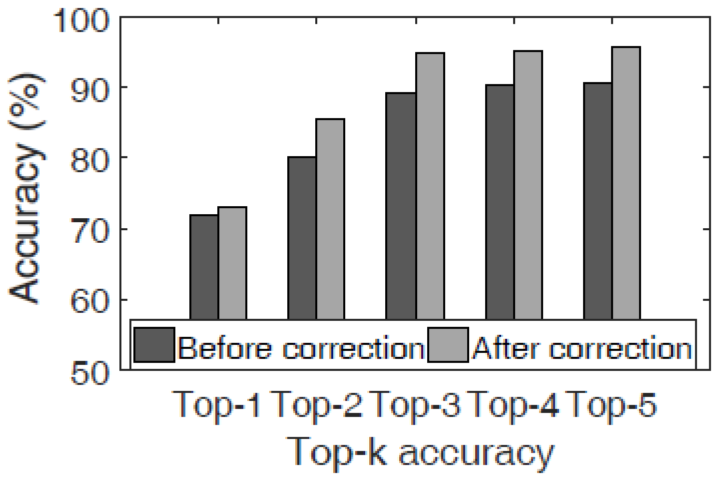
- 94.4%, 94.9% and 93.2% in meeting room, lab area and entertainment zone.

Different participants

- 95.6%, 93.5%, 93.1%, 93.0%, 94.8% and 95%, respectively.
- The standard deviation is about 1.1%.

Evaluation – word recognition





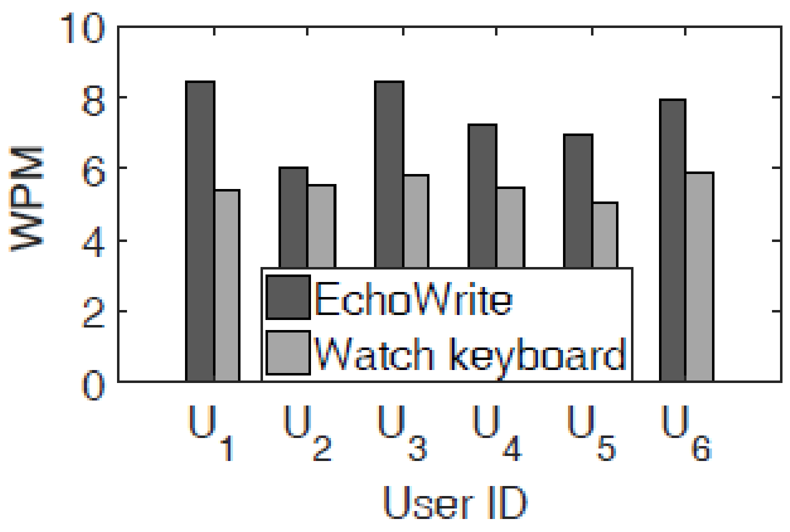
Different words

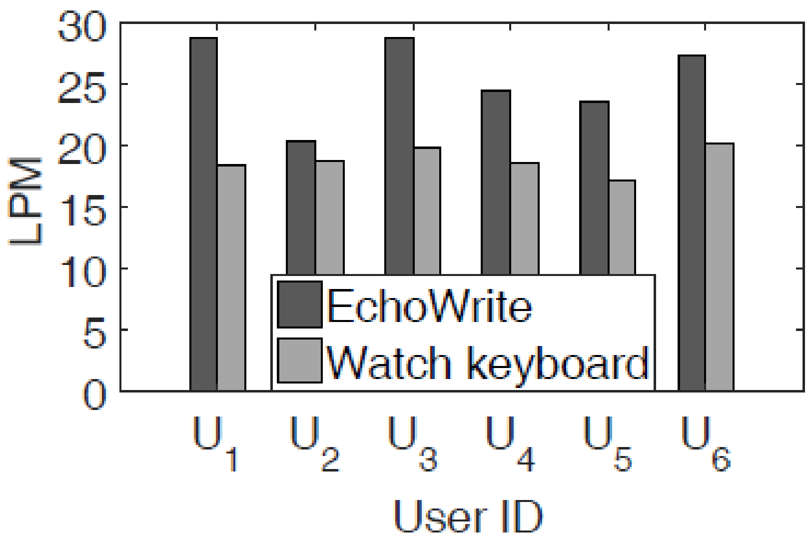
- Average top k accuracies over different words are 73.2%, 85.4%, 94.9%, 95.1% and 95.7%.
- Providing three candidates, the inferring accuracy is up to 94.9%.

Linguistic model

- The average accuracies are 84.5% and 88.9%, for cases with and without Linguistic model.

Evaluation – input speed



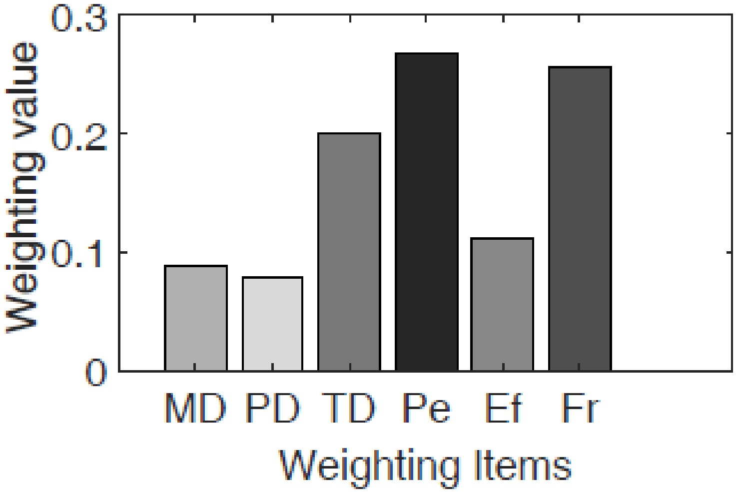


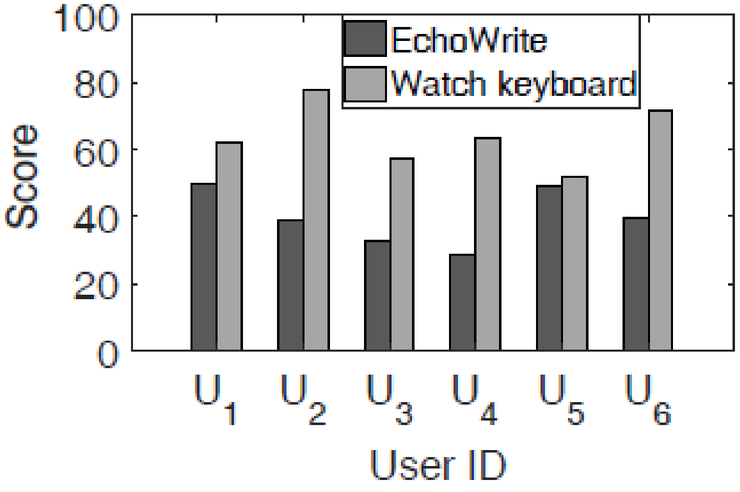
***WPM**: Words Per Min* ***LPM**: Letters Per Min*

Speed of text-entry

- The average texts-entry speeds over all participant with *EchoWrite* and smartwatch keyboard are 7.5 WPM and 5.5 WPM, respectively.

Evaluation - User experience





NASA-TLX workload factors:

- *mental demand (MD)*
- *physical demand (PD)*
- *temporal demand (TD)*
- *performance(Pe)*
- *effort (Ef)*
- *frustration (Fr)*

User experience assessment

- Performance, Frustration and temporal demand are the top 3 factors that affect users' assessment on a text-input system.
- the overall score of participants shows, smartwatch soft keyboard has higher workload than EchoWrite.

EchoWrite – conclusion

We design and implement EchoWrite which can recognize fine-grained finger-writing strokes without training.

With high-frequency signals and careful-designed processing pipeline, EchoWrite is robust to ambient noise and moving interference.

EchoWrite enables users to perform text-entry in the air with the speed of 7.5 WPM with the commodity microphone and speaker.

THANKS

Qiang Yang

yangqiang2016@email.szu.edu.cn

https://sites.google.com/view/yang-qiang/

Shenzhen University