Push the Limit of Acoustic Gesture Recognition

Yanwen Wang, Jiaxing Shen, Yuanqing Zheng

The Hong Kong Polytechnic University, Hong Kong, China

INFOCOM 2020, Online
July 2020

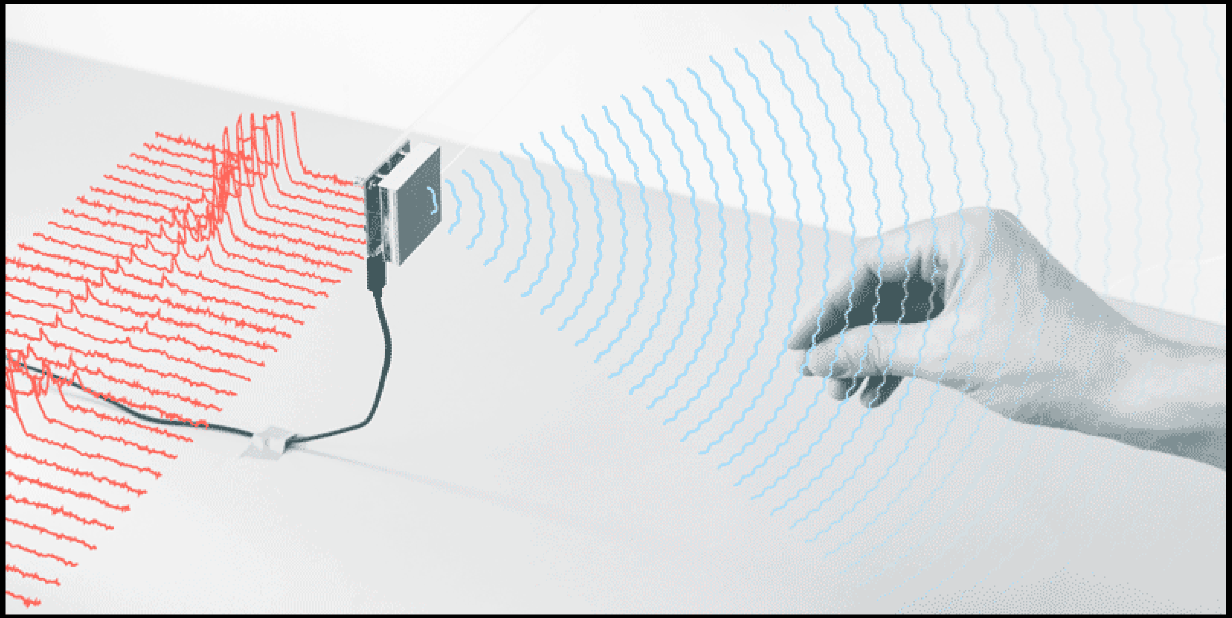
Contact-free Gesture Recognition

Camera-based

Camera-based

- ➢ Rely on Line-of-Sight path
- ➢ Rely on good lighting conditions

Contact-free Gesture Recognition



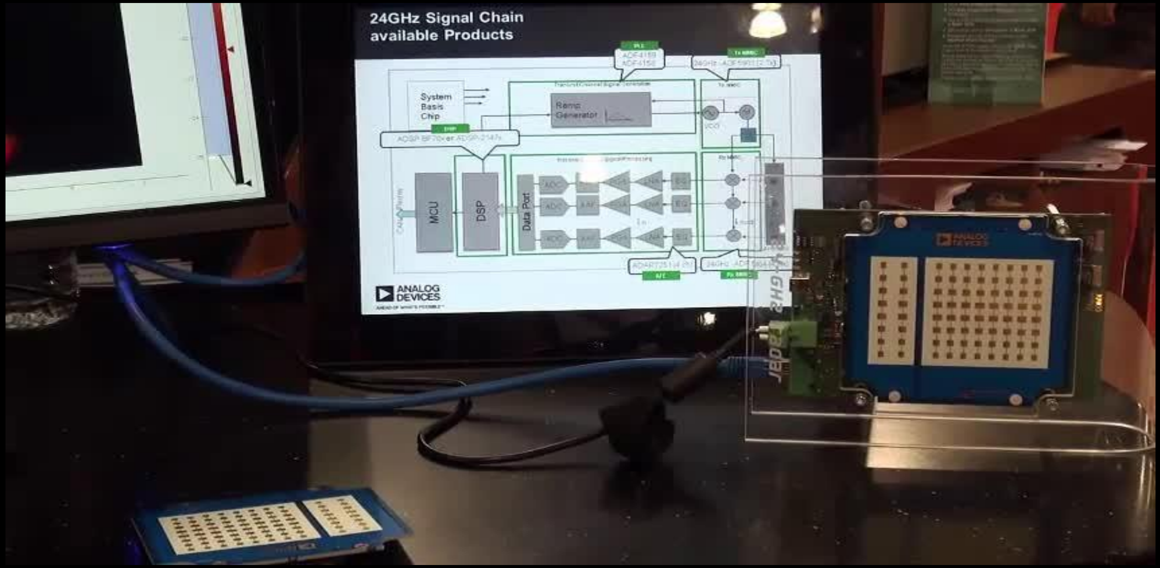
Millimeter Wave

Millimeter Wave

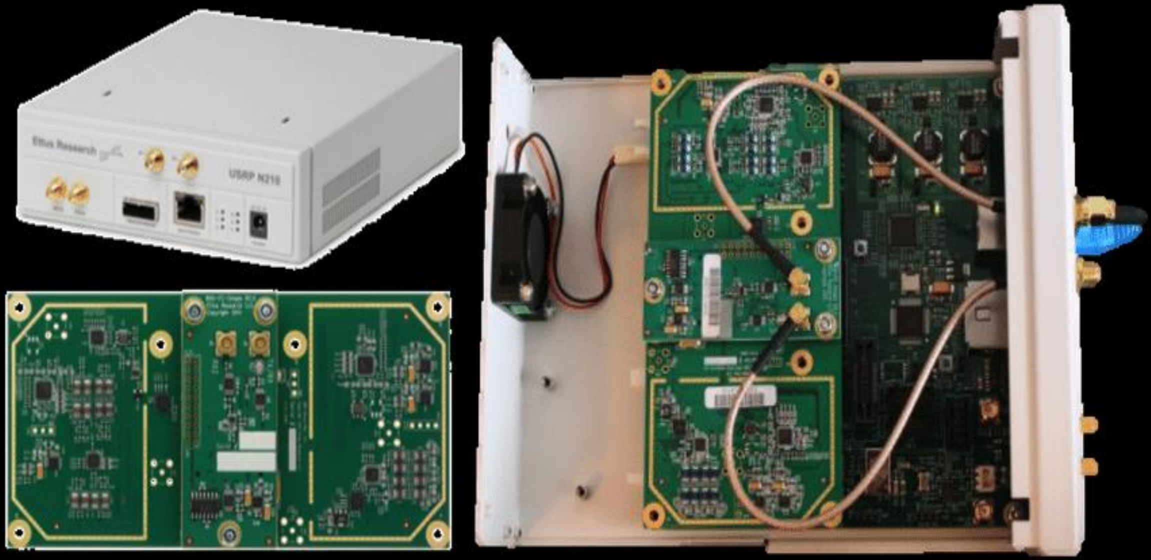
https://atap.google.com/soli/

- Works in 60 GHz frequency band
- Not allowed in some countries

Contact-free Gesture Recognition



FMCW



USRP

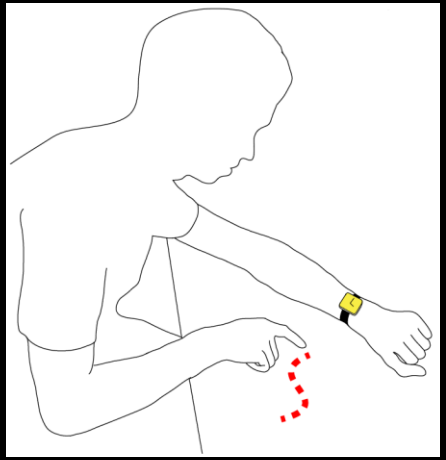
[1] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013. Whole-home gesture recognition using wireless signals. In ACM MobiCom 2013.
[2] Kellogg, Bryce, Vamsi Talla, and Shyamnath Gollakota. 2014. Bringing gesture recognition to all devices. In USENIX NSDI 2014.

- **Require specialized devices**
- **High deployment costs**

Benefit from Gesture Recognition

Avoid contact contamination in public area

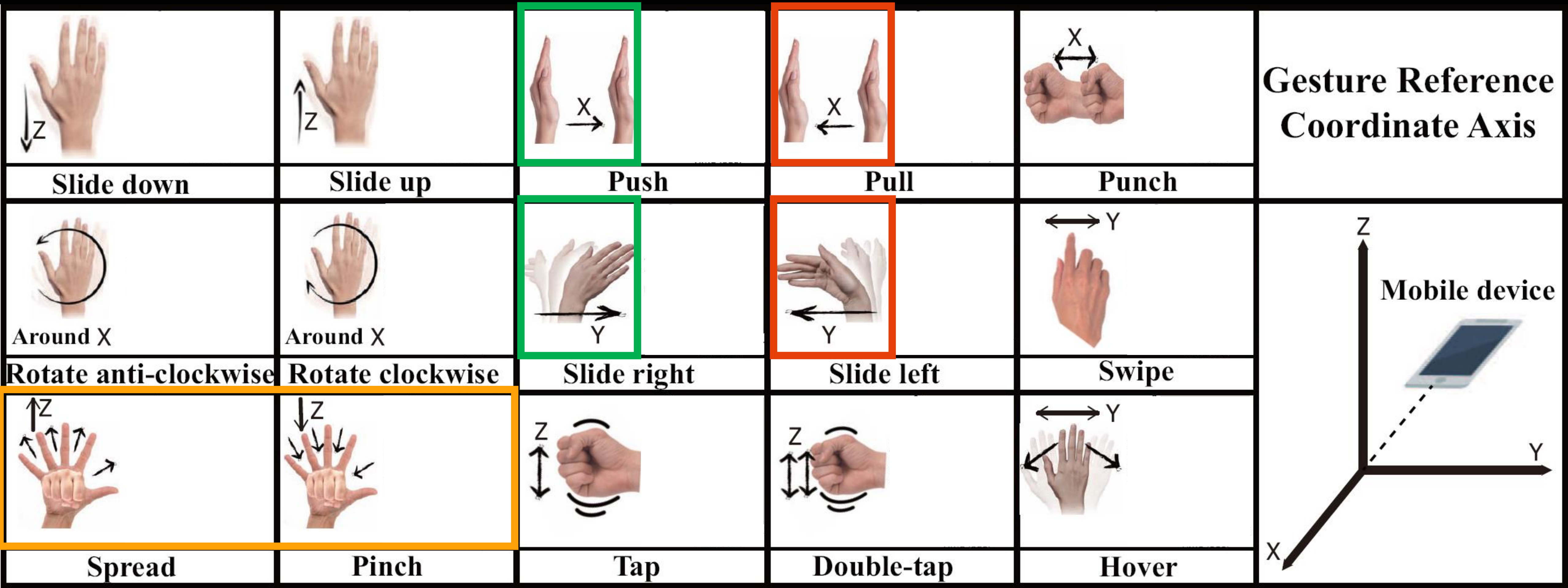
Acoustic-based Finger Tracking



[1] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota. Fingerio: Using active sonar for fine-grained finger tracking. In ACM CHI, 2016.
[2] W. Wang, A. X. Liu, and K. Sun. Device-free gesture tracking using acoustic signals. In ACM MobiCom, 2016.
[3] S. Yun, Y.-C. Chen, H. Zheng, L. Qiu, and W. Mao. Strata: Fine-grained acoustic-based device-free tracking. In ACM MobiSys, 2017.

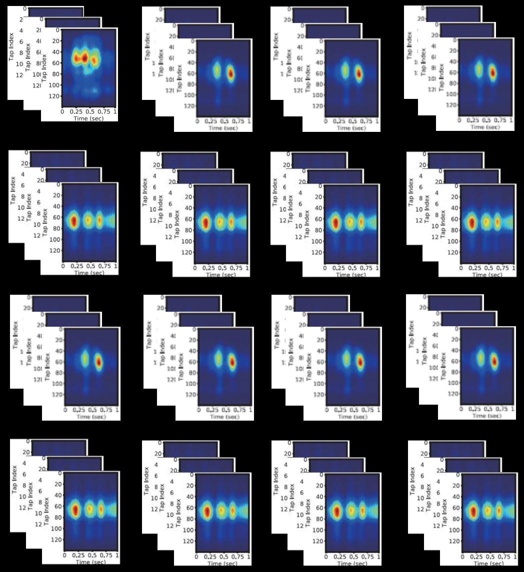
- ➢ **Model the whole finger/hand as a single reflection point**
- ➢ **Insufficient resolution for gesture recognition**

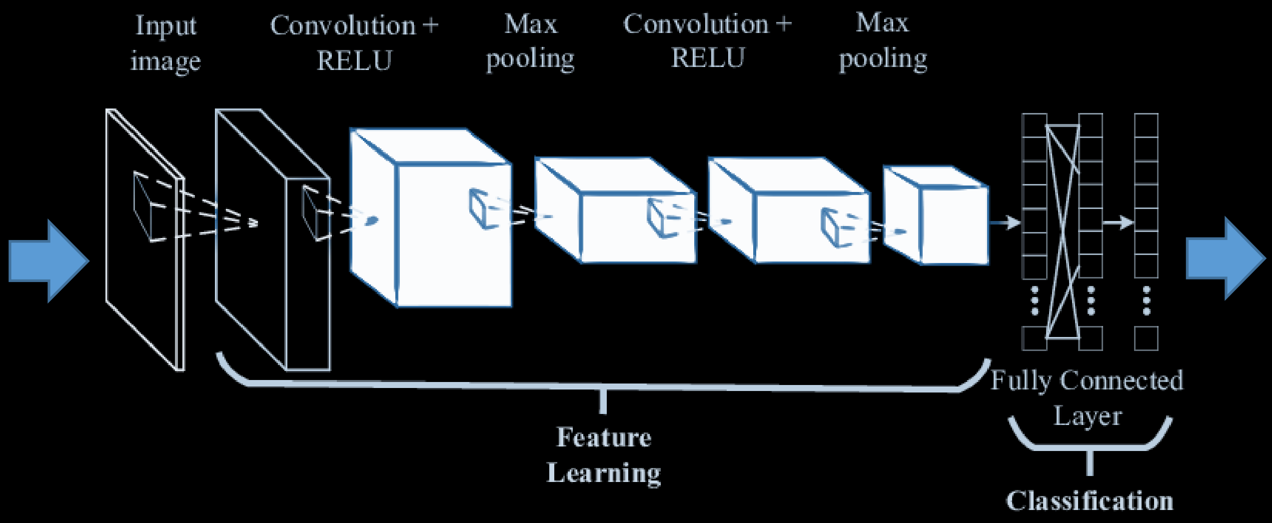
RobuCIR

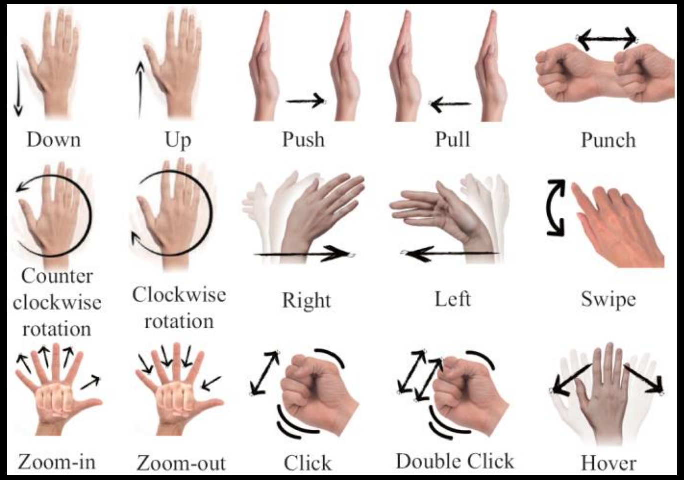


Acoustic signal based gesture recognition system with high robustness

Use Neural Network

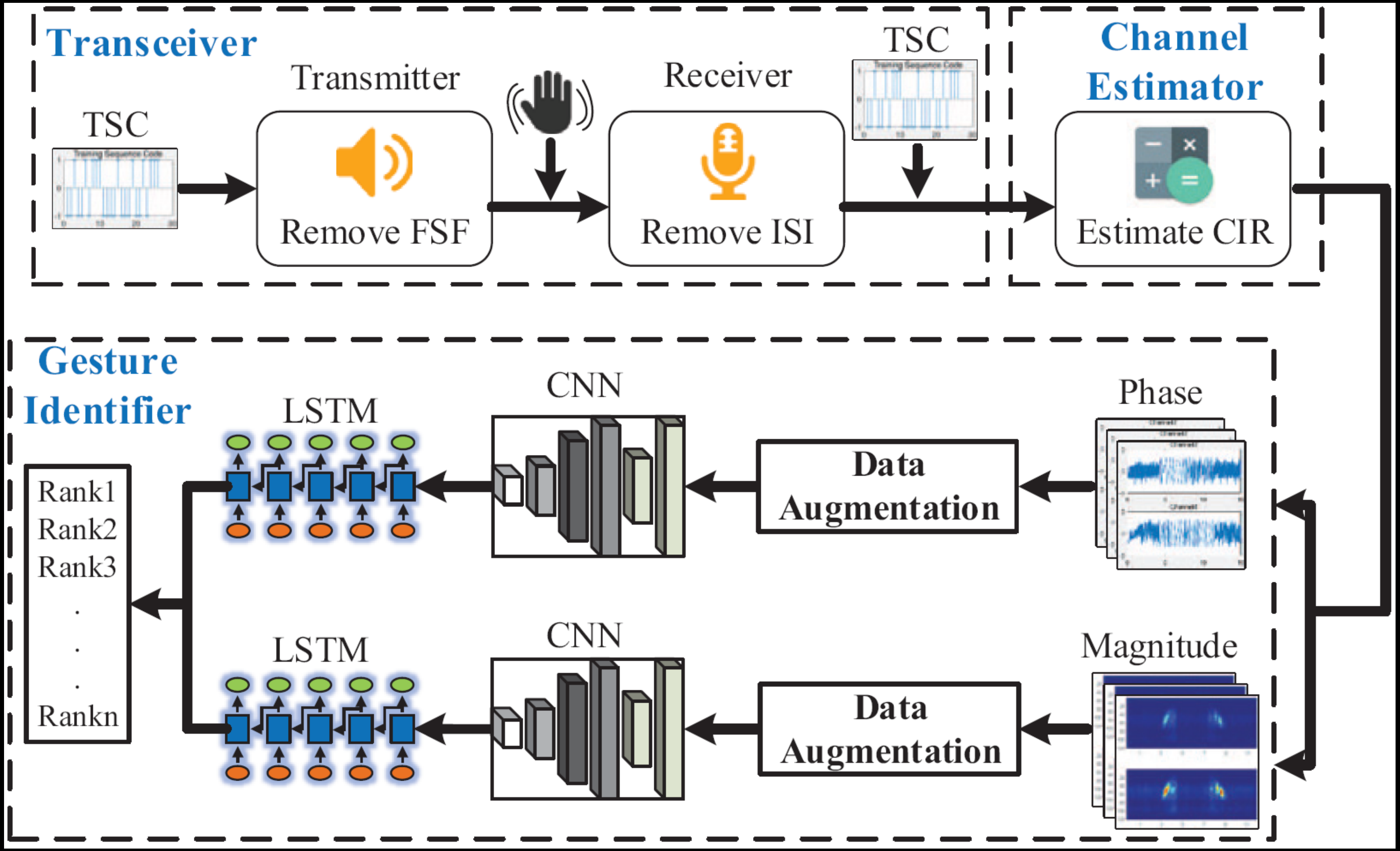




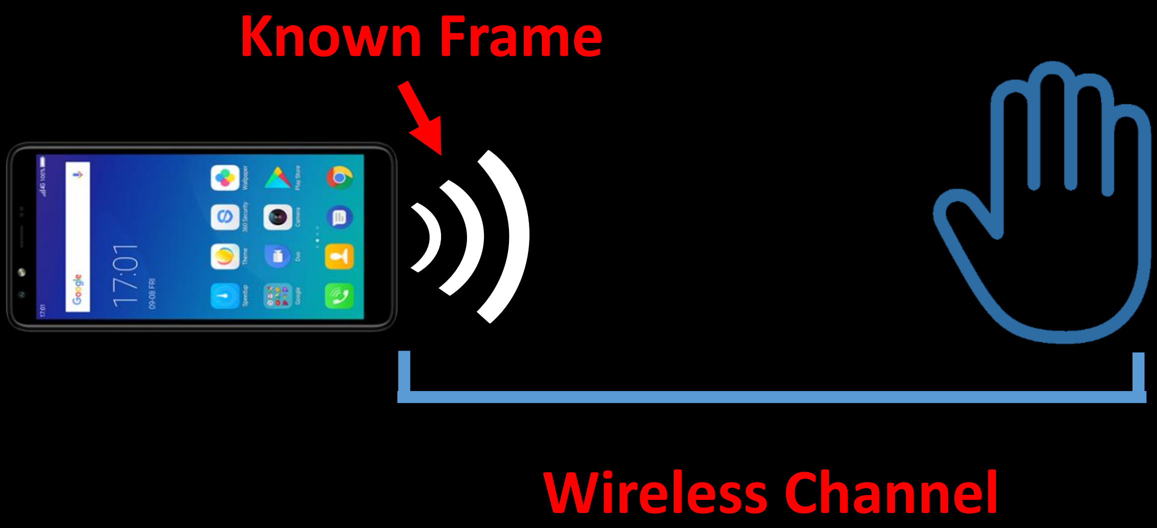


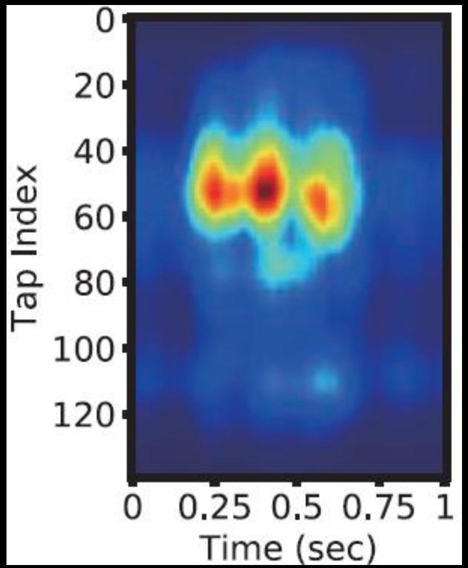
- ➢ **Insufficient data for training**
- ➢ **Low robustness**

Overview


Transceiver
Transmitter
TSC
Remove FSF
Receiver
Remove ISI
TSC
Channel Estimator
Estimate CIR
Gesture Identifier
LSTM
CNN
Phase
Data Augmentation
Rank1
Rank2
Rank3
Rankn
LSTM
CNN
Magnitude
Data Augmentation

Measure the Channel: Channel Impulse Response





Least Square (LS) channel estimation

$$h = arg\min_{h} \|r - Mh\|^2$$

M is the training matrix consisting of transmitted circulant orthogonal codes

r is the received signal

Each value in h measures the channel information of a certain propagation delay range.

Limitation of Single Frequency

❑ Frequency Selective Fading

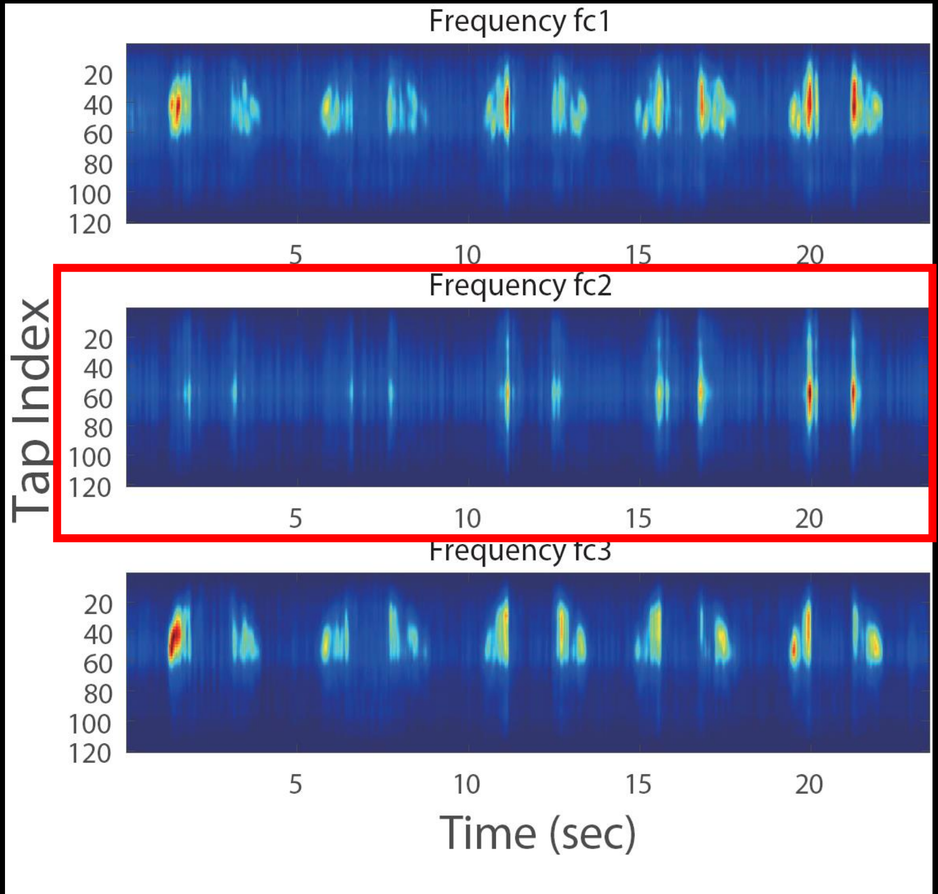
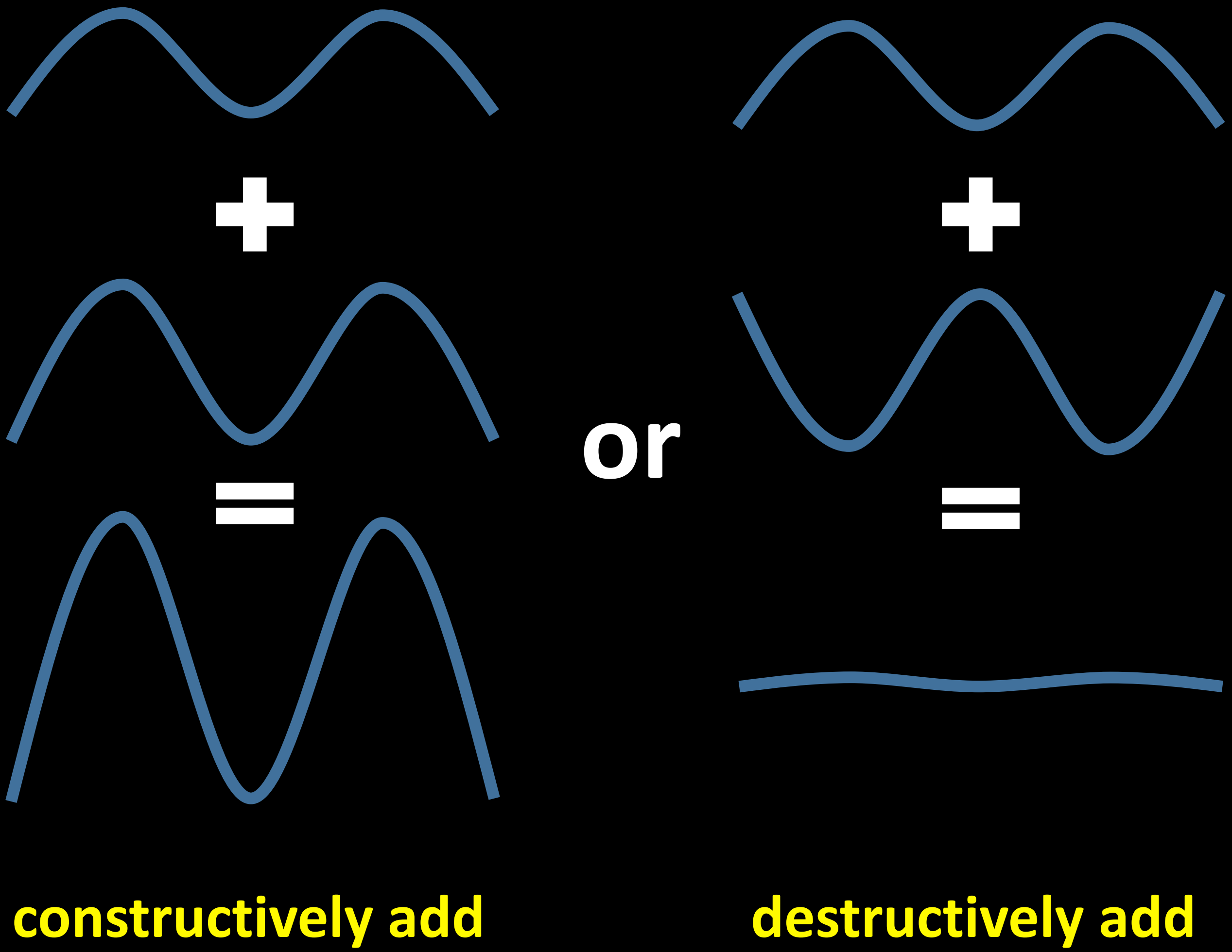
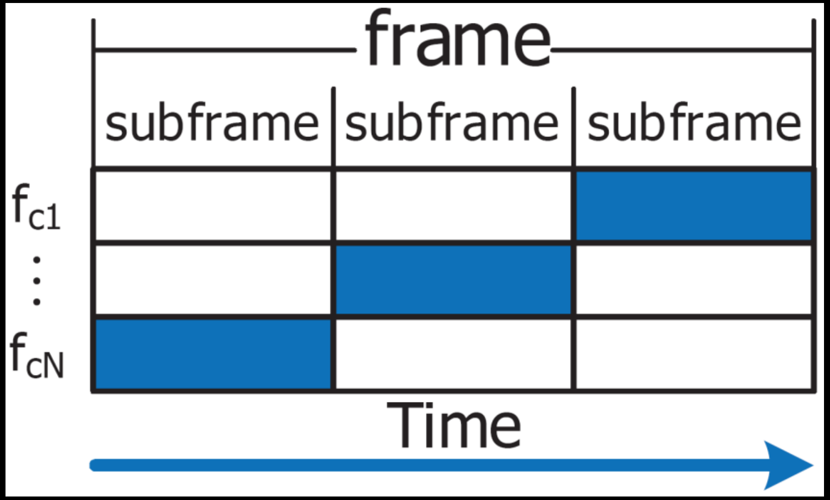


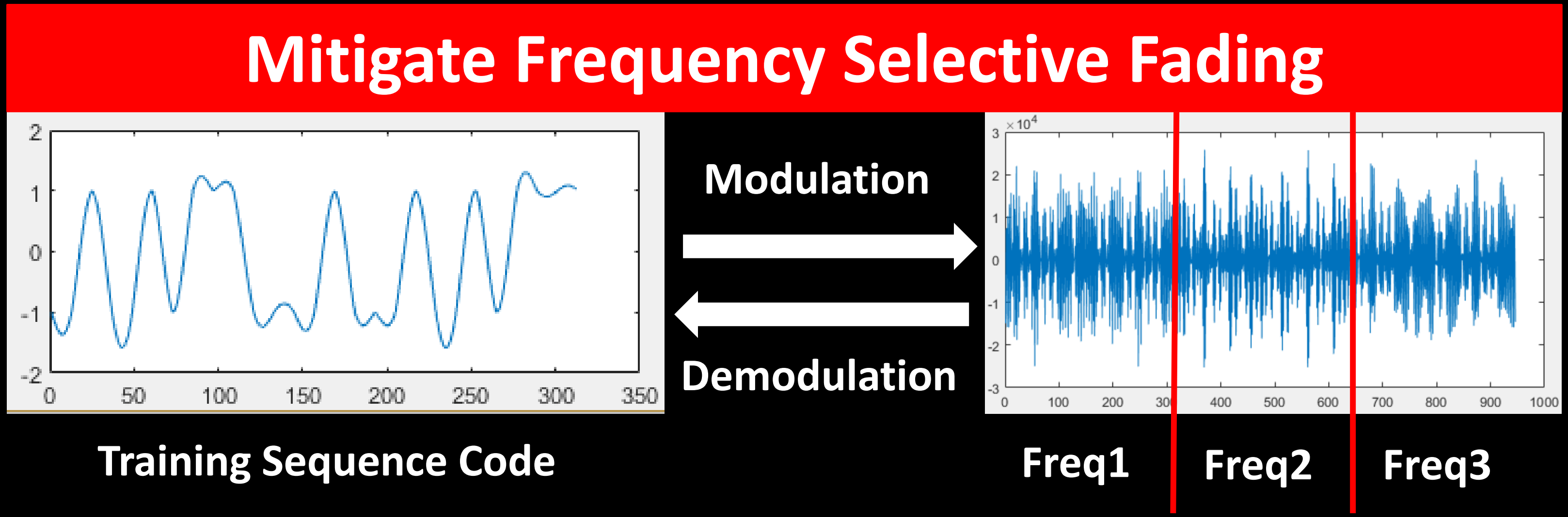
Fig. 4. CIR when performing push and pull.



Our Solution

❑ Frequency Hopping Scheme





Demodulated baseband signal involves information from multiple frequencies

Apply Data Augmentation to Raw CIR Measurement

① Different speeds

② Different hand-smartphone distances

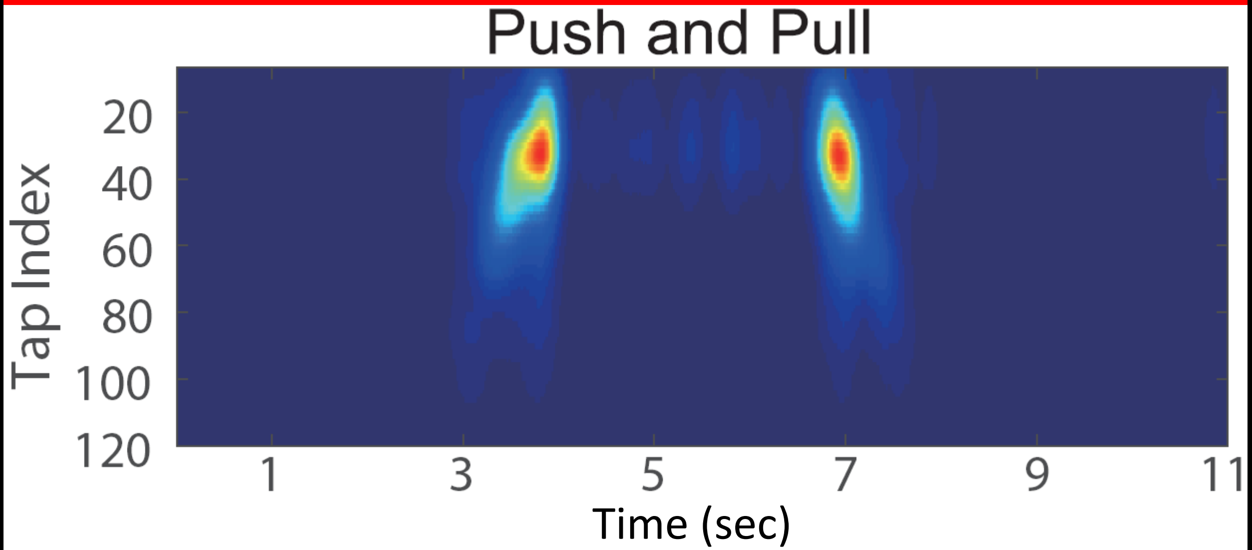
③ Blockage of LoS

④ Noisy environments

⑤ Different angles to smartphone

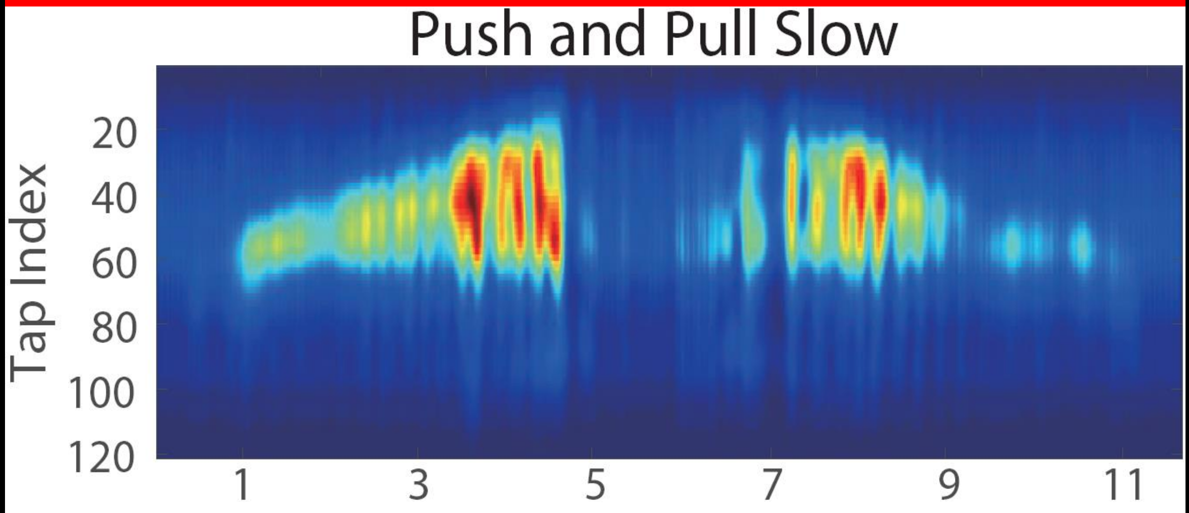
Impact Factor: ① Different speeds of the gestures

Reference Gesture



Push and pull at 0 ~ 20cm

Different speed of the gesture



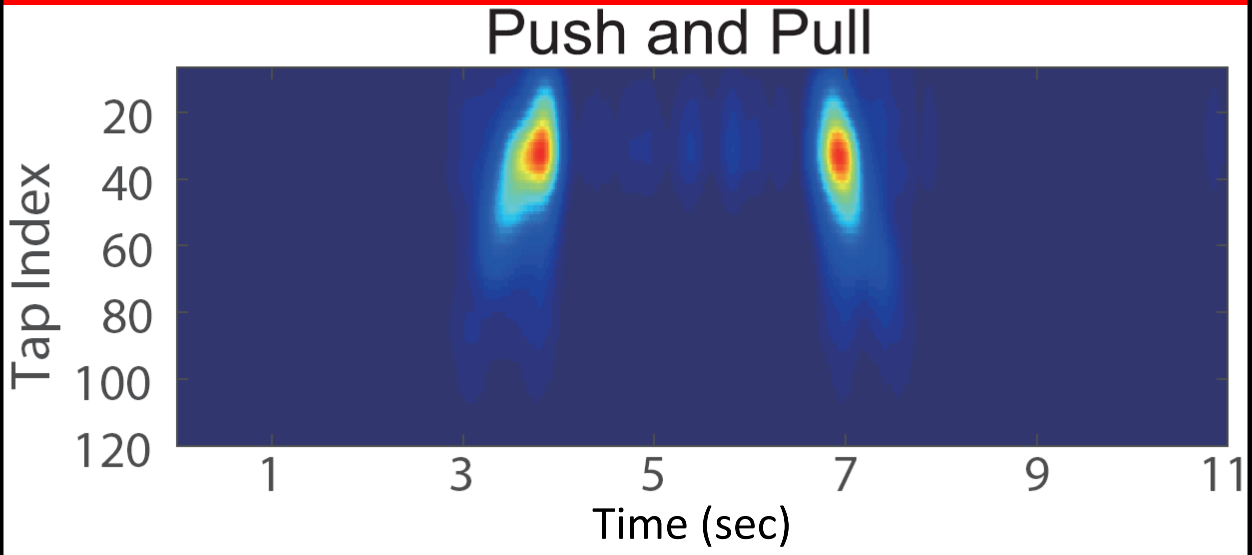
Push and pull at slow speed

Data augmentation: Horizontally expanding or contracting an original CIR measurement

02 Methodology

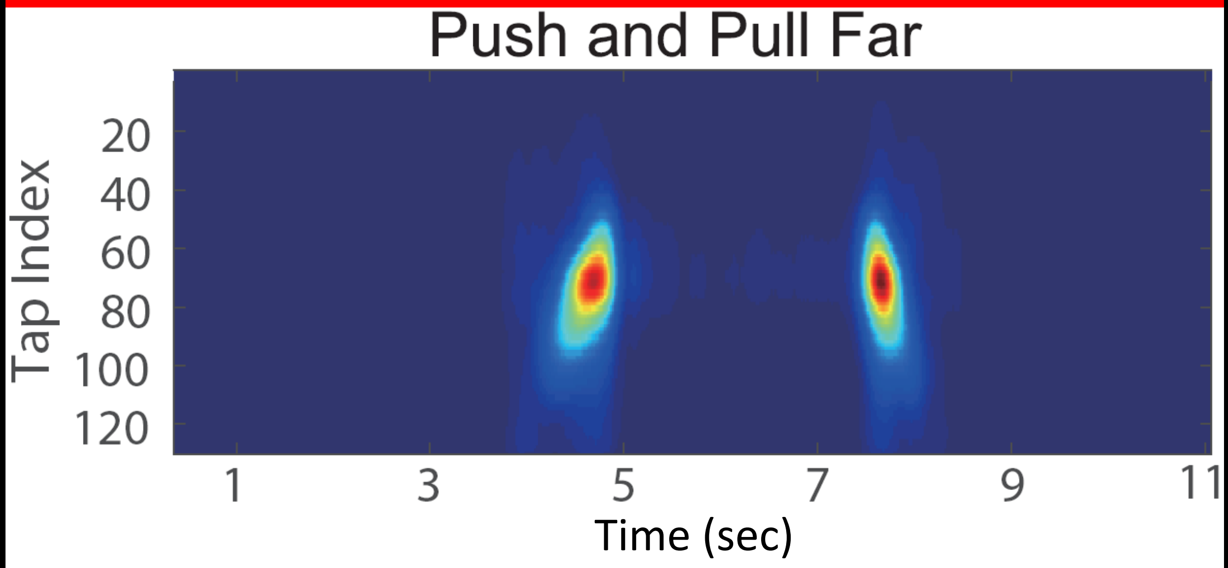
Impact Factor: ② Different distances to the receiver

Reference Gesture



Push and pull at 0 ~ 20cm

Different distances to the receiver

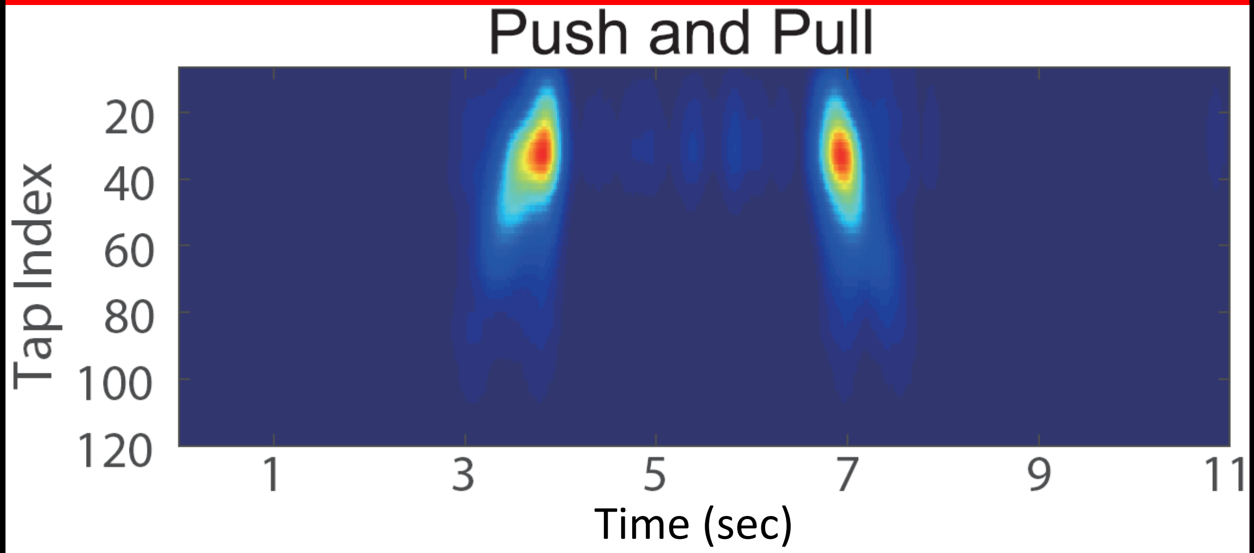


Push and pull at 20 ~ 40cm

Data augmentation: Vertically drifting in tap indexes

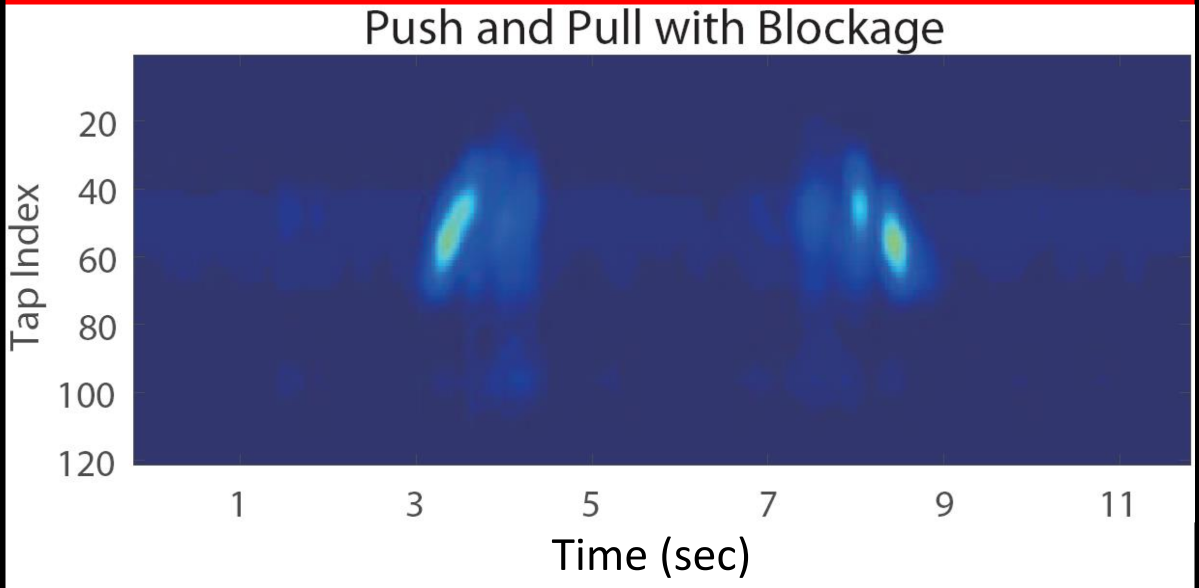
Impact Factor: ③ Blockage of transceiver

Reference Gesture



Push and pull at 0 ~ 20cm

Blockage of transceiver

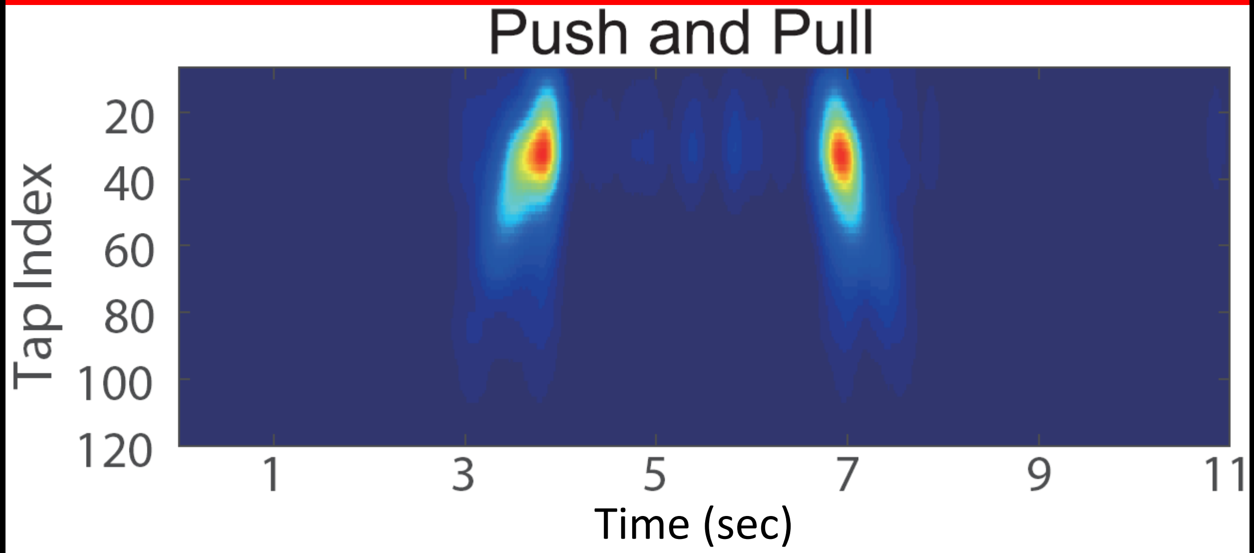


Push and pull in the pocket

Data augmentation: Scale and normalize the CIR magnitude measurements

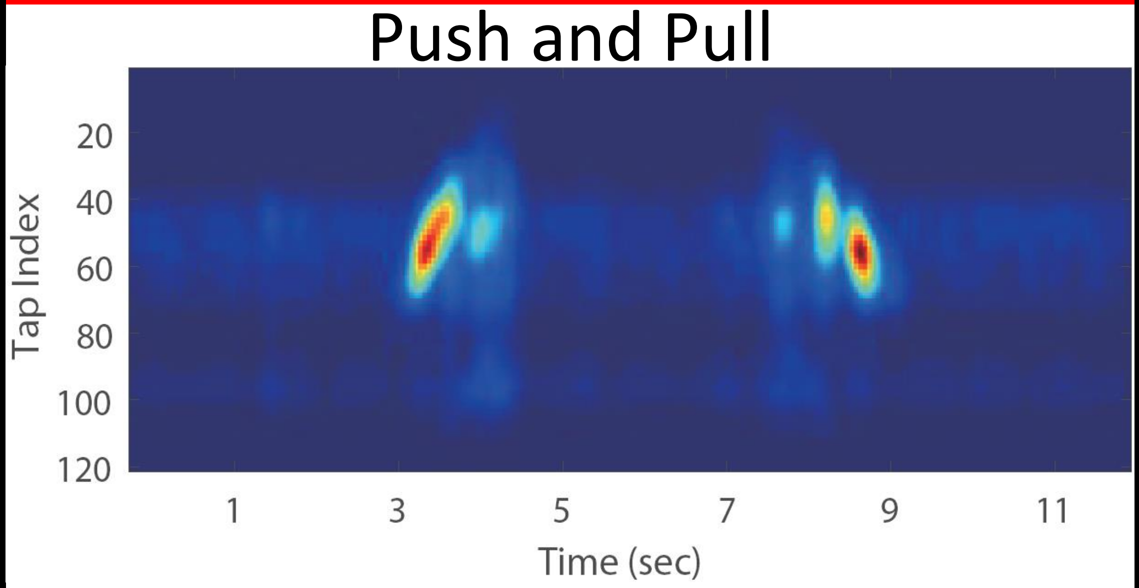
Impact Factor: ④ Noisy environments

Reference Gesture



Push and pull at 0 ~ 20cm

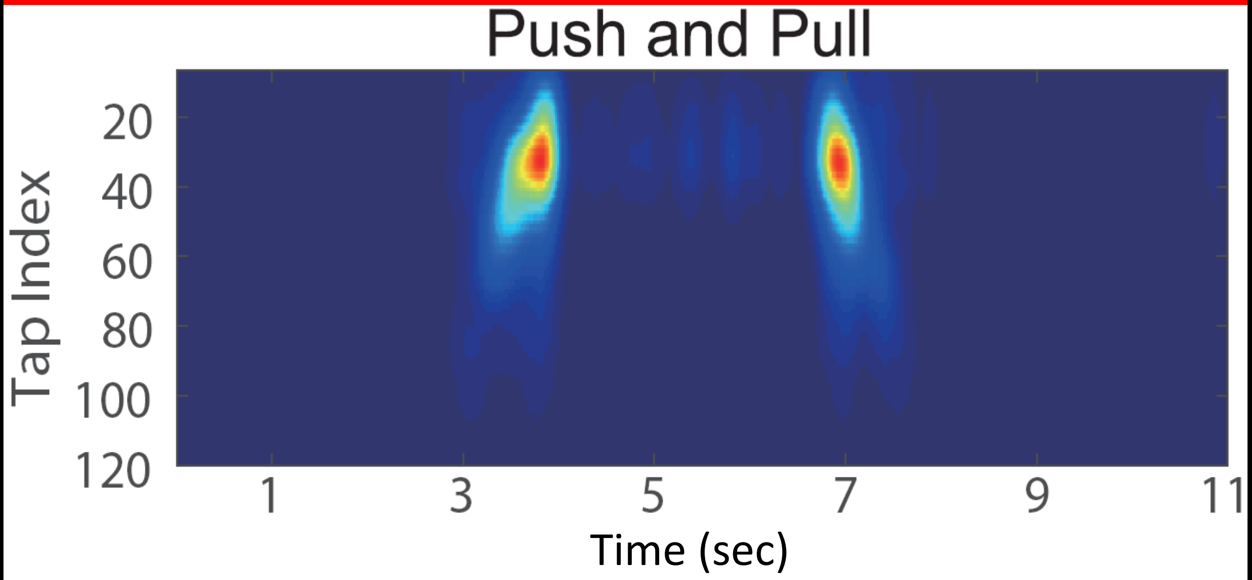
Noisy environment



Push and pull near the loudspeaker

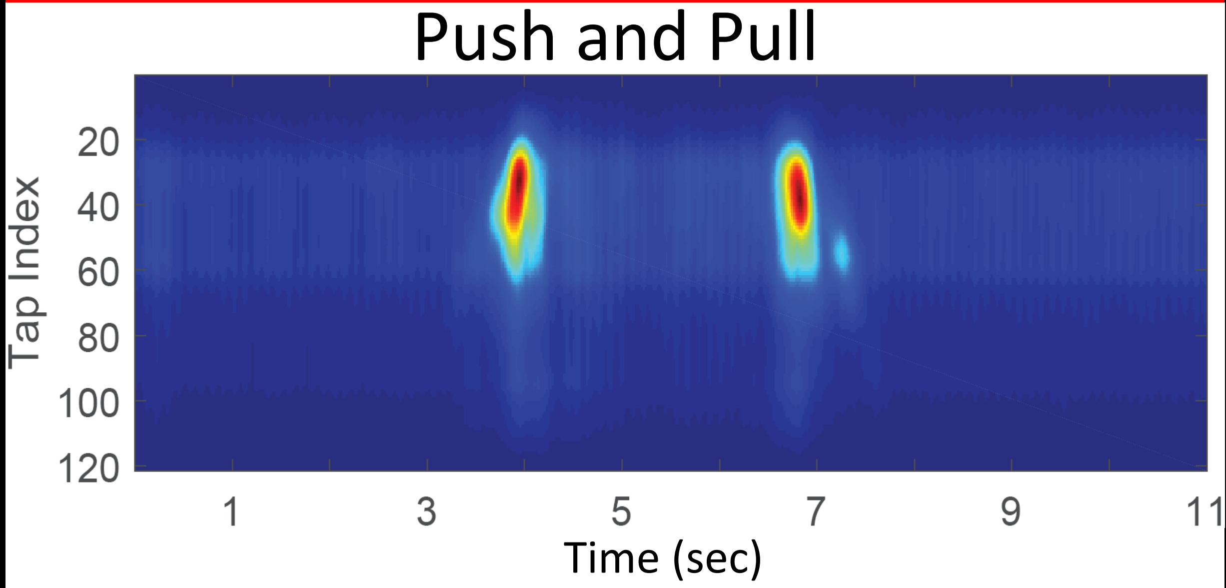
Impact Factor: ⑤ Different angles to device

Reference Gesture



Push and pull at 0 ~ 20cm

Different angles to device



Push and pull at different angles to device

System Performance

Robustness

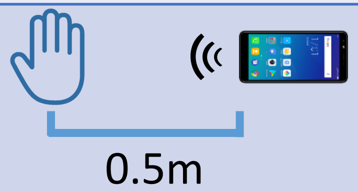
Under Different Experiment Settings

Experiment Settings

Transmitted Frequency	18KHz, 20KHz, 22KHz
Transmitted Baseband Signal	26-bit TSC
Smartphone	Samsung S9 Plus, Samsung S7 Edge, Google NEXUS 5
Eight Volunteers	5 males and 3 females
Environment	8 different rooms with different layouts, 5 usage scenarios
Collected Data	3600 samples
Benchmark	UltraGesture[1]
Data Augmentation Rate	100X
Neural Network	3-layer CNN + 8-cell 1 layer LSTM

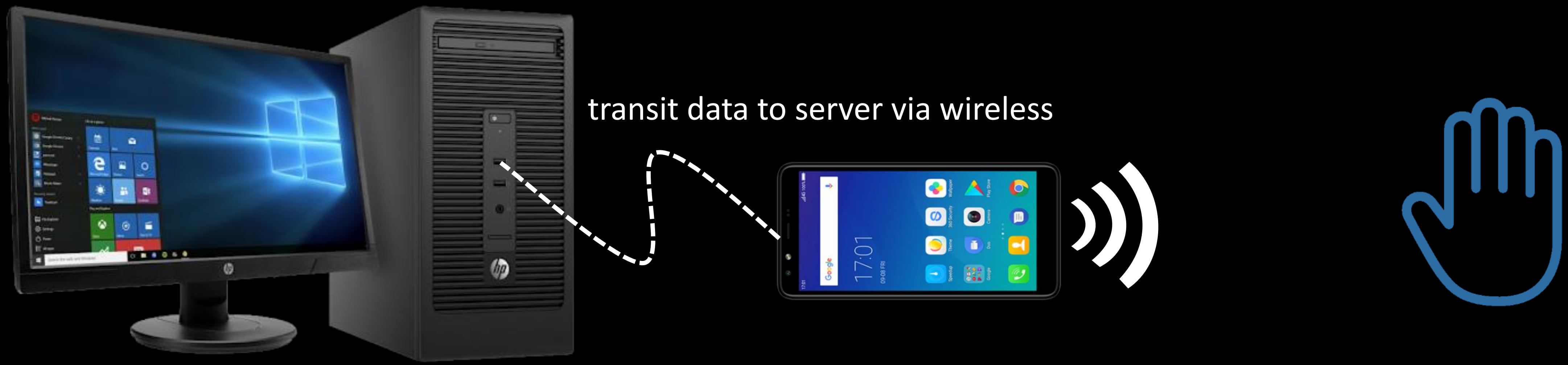
[1] Ling, Kang, Haipeng Dai, Yuntang Liu, and Alex X. Liu. Ultragesture: Fine-grained gesture sensing and recognition. In *IEEE SECON*, 2018.

Scenarios Settings

Different Speeds	At most 5$\times$(e.g., from 0.4s to 2s)
Different Distances	Up to 0.5m from smartphone 
Blockage of LoS	In a cotton bag
Different Angles	Ranging from $0^{\circ} \sim 180^{\circ}$ 
Noisy Environments	Another mobile phone plays music nearby 
Different Environments	$10 \times 8 \times 3m^3$ ~ $4 \times 4 \times 3m^3$ with different layouts

Model Training and Gesture Identification

- 10-fold cross-validation ---- train 6 & test 2
- Model size: 5.5M



- Intel(R) Xeon(R) E5-2620 v4 CPU @2.10GHz
- 32GB memory
- Two Nvidia GTX 1080 Ti GPU graphics cards.
- Using TensorFlow

CIR Measurements Calculation			Gesture Recognition
Frame detection	Down-conversion	LS	Coupled NN model
1.3ms	2.2ms	4.8ms	23ms

System Performance

Accuracy

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Slide down (1)	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Slide up (2)	0.00	0.98	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00
Push (3)	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pull (4)	0.01	0.00	0.00	0.97	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00
Punch (5)	0.00	0.01	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R-clock (6)	0.00	0.00	0.01	0.00	0.00	0.97	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01
R-anticlock (7)	0.00	0.00	0.01	0.00	0.00	0.00	0.96	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.00
Slide right (8)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.01	0.01	0.00	0.00
Slide left (9)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Swipe (10)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
Spread (11)	0.01	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.95	0.00	0.01	0.00	0.01
Pinch (12)	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.97	0.01	0.00	0.01
Tap (13)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Double-tap (14)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.99	0.00
Hover (15)	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.99

True Activities

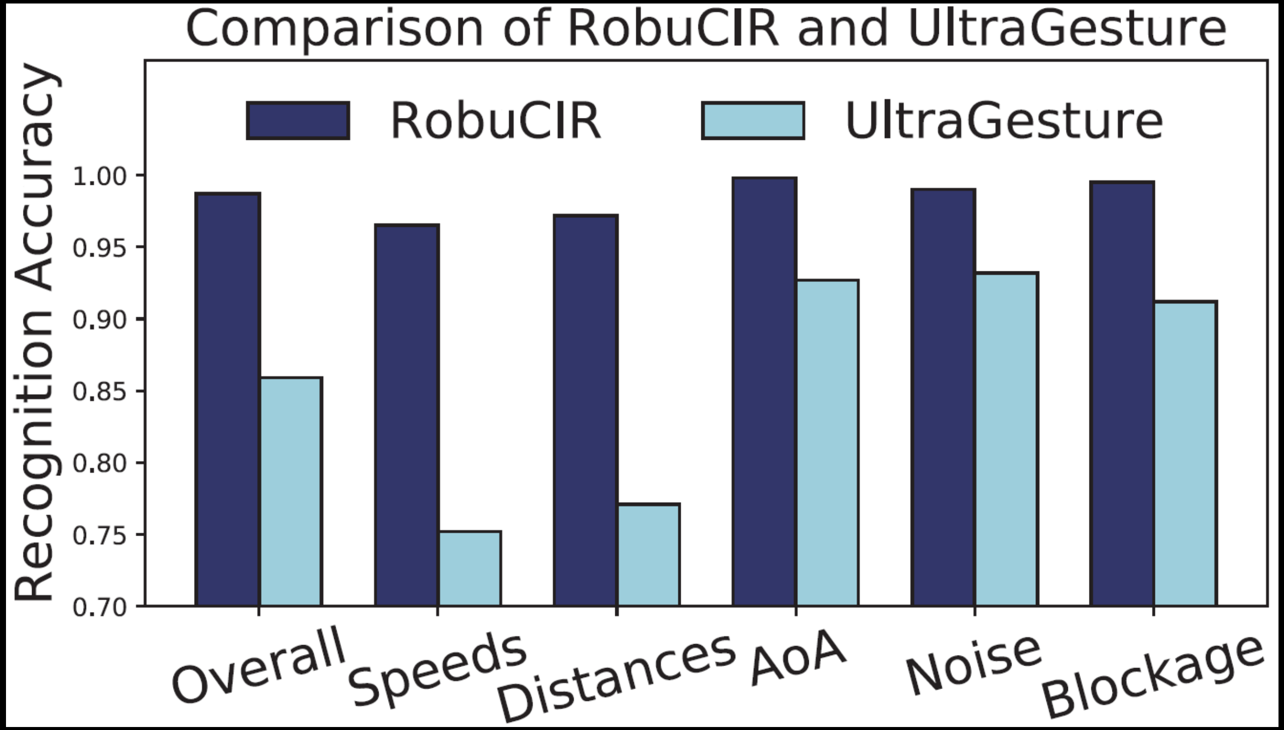
Overall Performance

98.4%

Improvement of Robustness

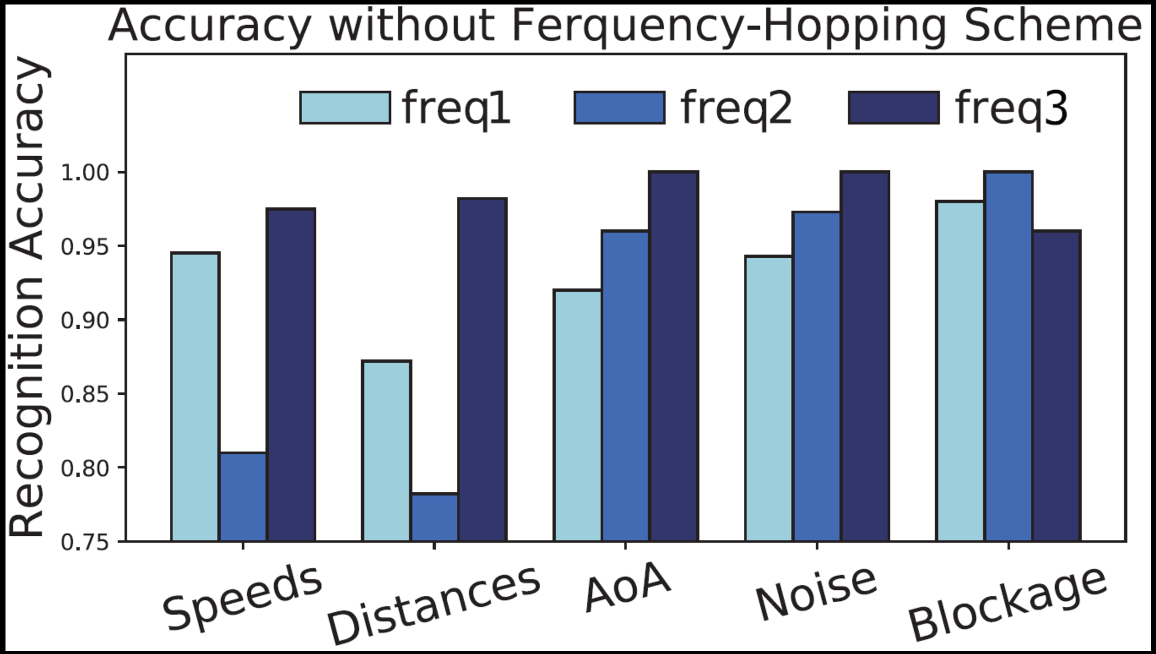
Robustness Improvement

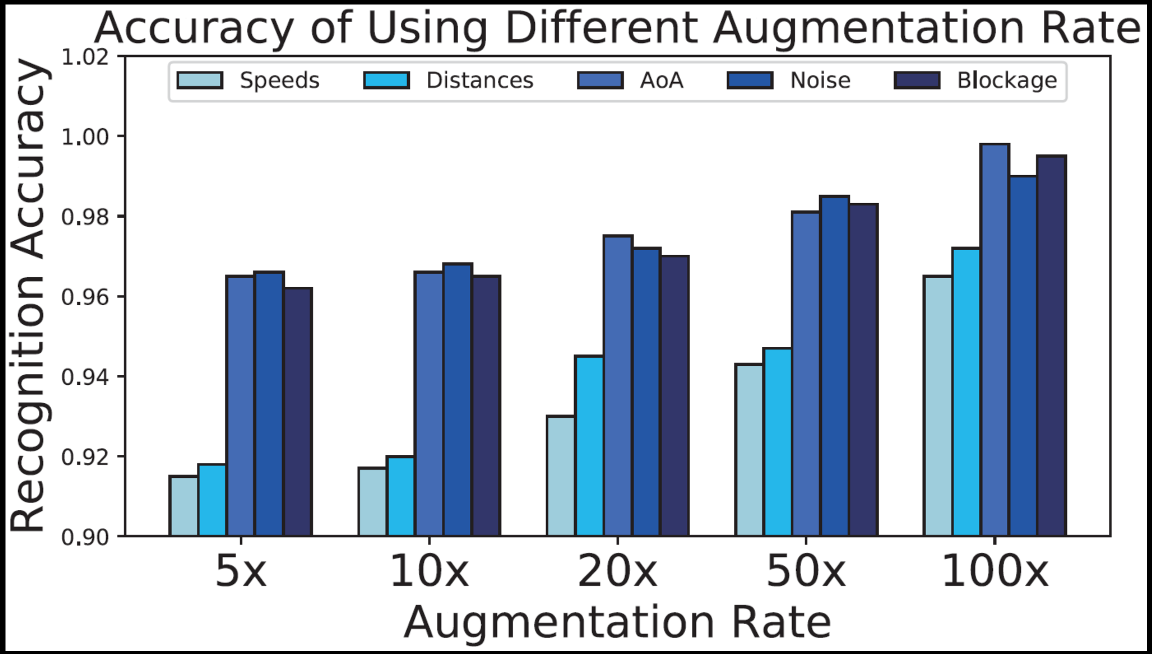
21%



UltraGesture does not consider gestures under different impact factors

Under Different Experiment Settings





- Using single frequency may decrease the system performance.
- A Larger augmentation rate covers more variations of the gestures.

Conclusion

04 Conclusion

- **Build an acoustic signal based system which can identify 15 types of gestures with high robustness and accuracy**
- **Adopt frequency hopping scheme to mitigate FSF**
- **Conduct data augmentation on raw CIR data to effectively train neural network models**
- **Outperform state-of-the-art work and achieve an overall accuracy of 98.4%.**

Thank you!

Email: yan-wen.wang@polyu.edu.hk
Homepage: https://www.comp.polyu.edu.hk/~csywwang

Yanwen Wang, Jiaxing Shen, Yuanqing Zheng

The Hong Kong Polytechnic University, Hong Kong, China

INFOCOM 2020, Online
July 2020