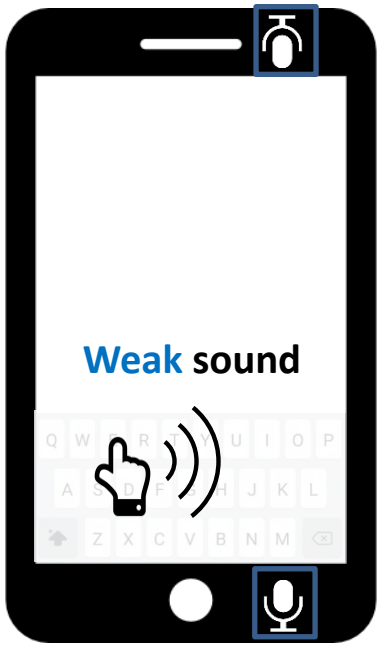
Mobile Phones Know Your Keystrokes through the Sounds from Finger's Tapping on the Screen

Zhen Xiao[1], Tao Chen[1], Yang Liu[1], Zhenjiang Li[1,2]
[1]Department of Computer Science, City University of Hong Kong
[2]City University of Hong Kong Shenzhen Research Institute

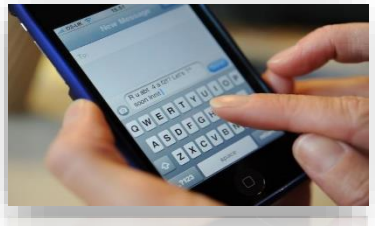
CityU
香港城市大學
City University of Hong Kong

Tapping sound on the mobile phone

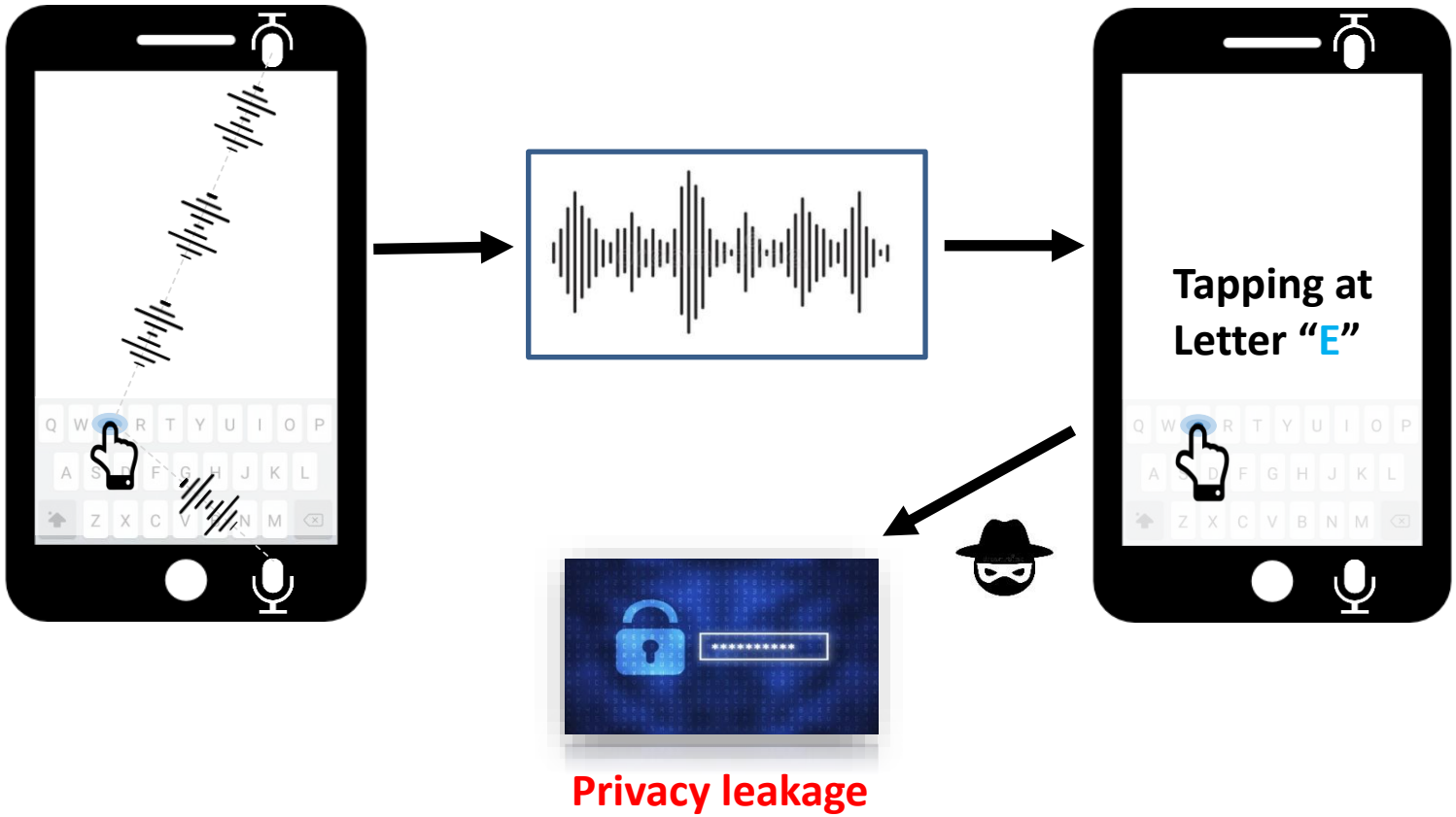


Privacy:

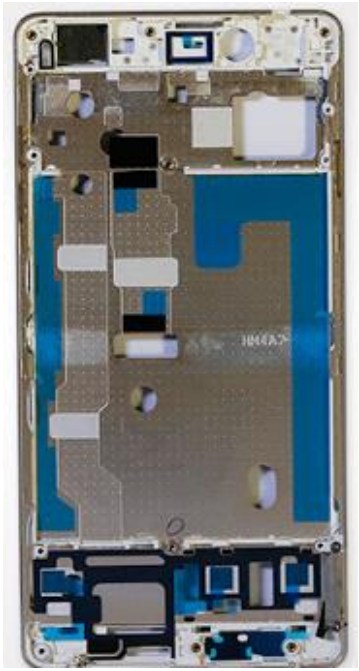
- **Message**
- **Password**
- **Bank account**



Infer keystrokes

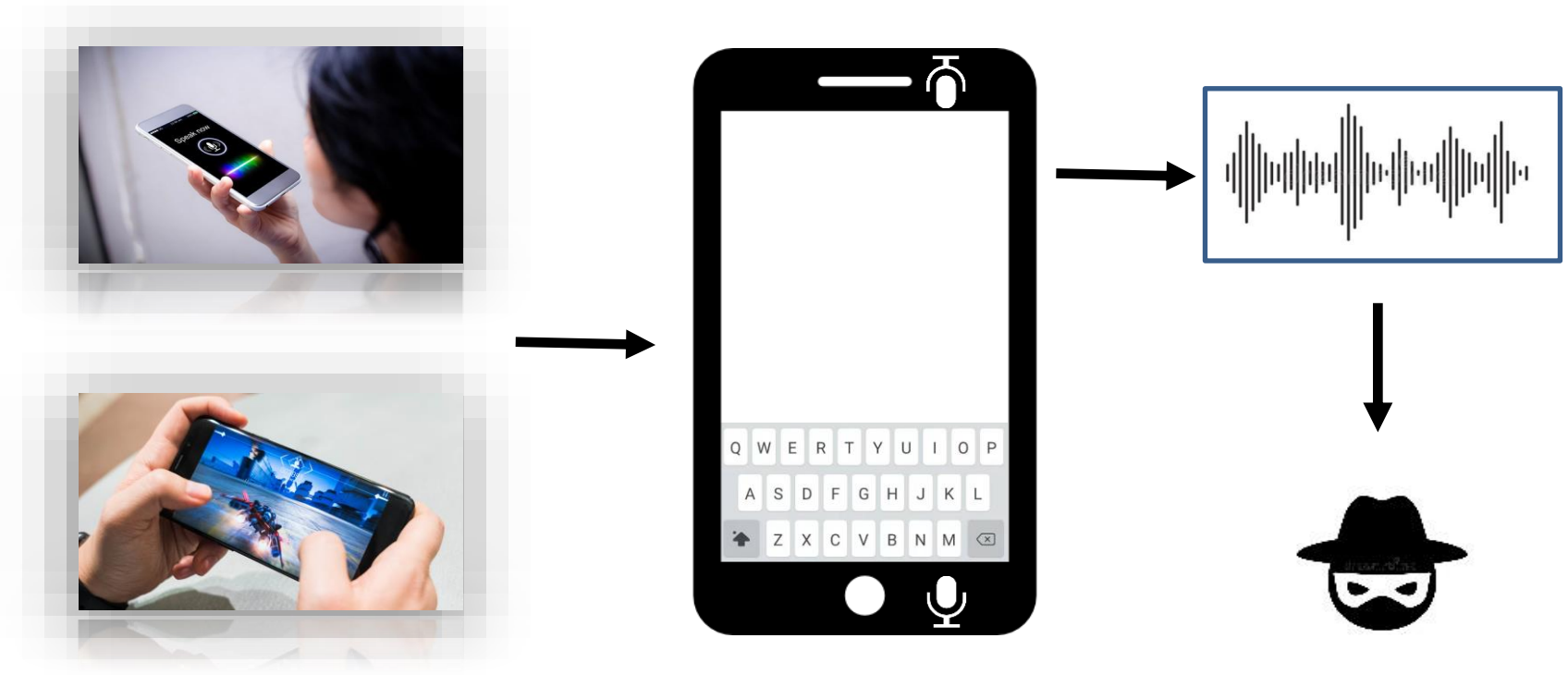


Distinct tapping sound

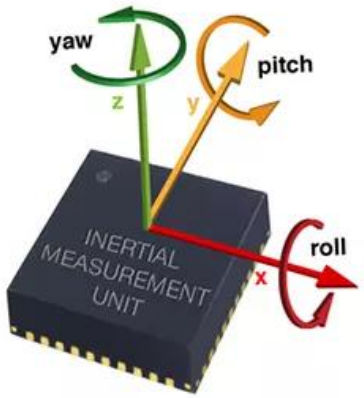


- Heterogeneous structure
- Distinct sound

Hacking acoustic signal



Existing methods



[WiSec,12]

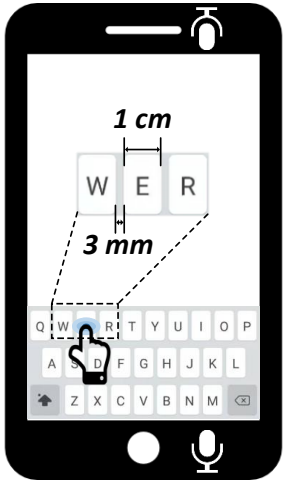
[WiSec,14]

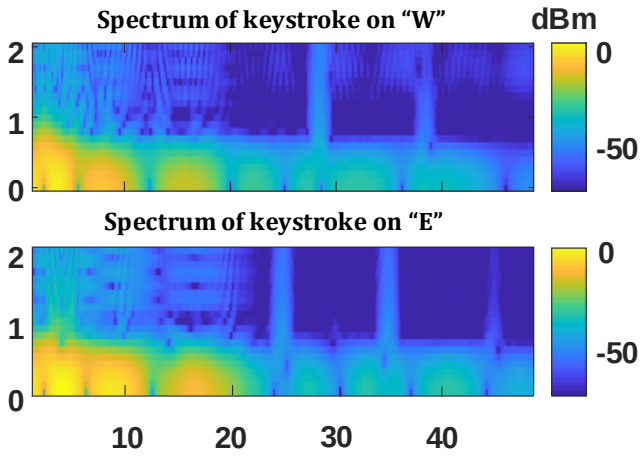
- Microphones and IMU sensors
- Fake keyboard

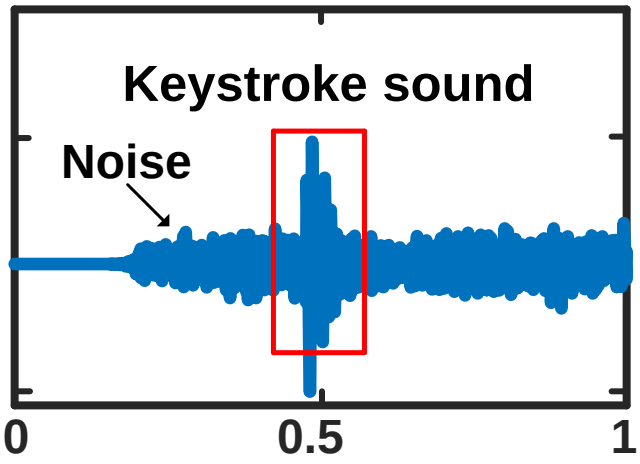
Stealthiness

Challenge-I: keystroke recognition

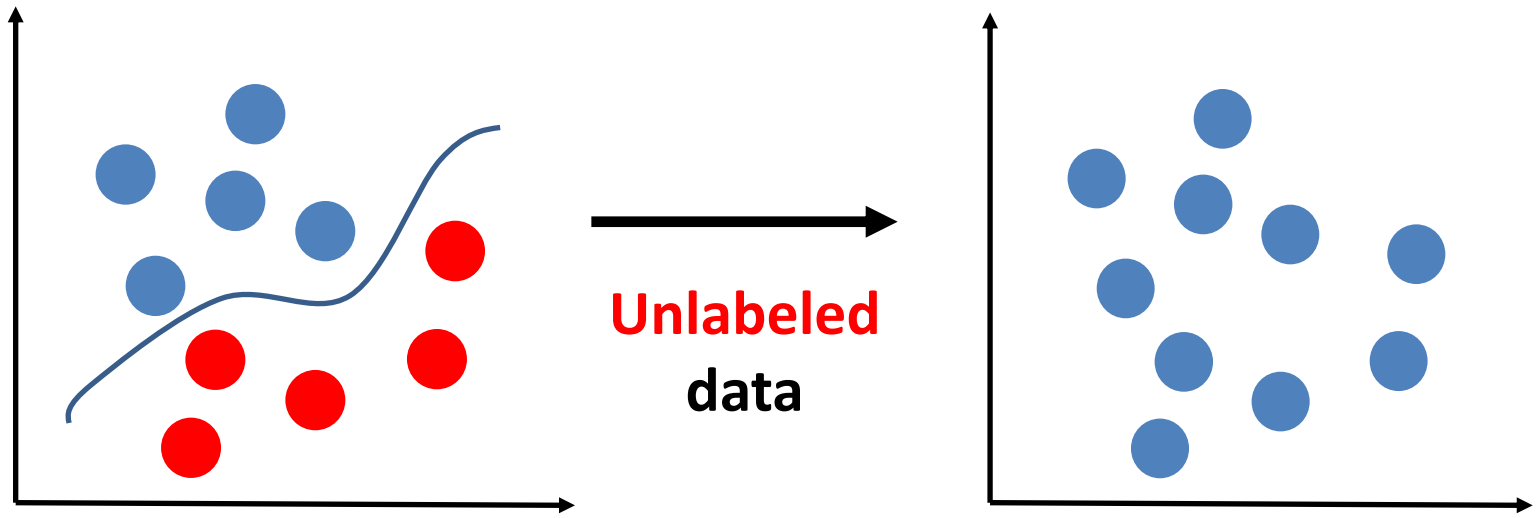
- **Difficult** task with **weak** signal





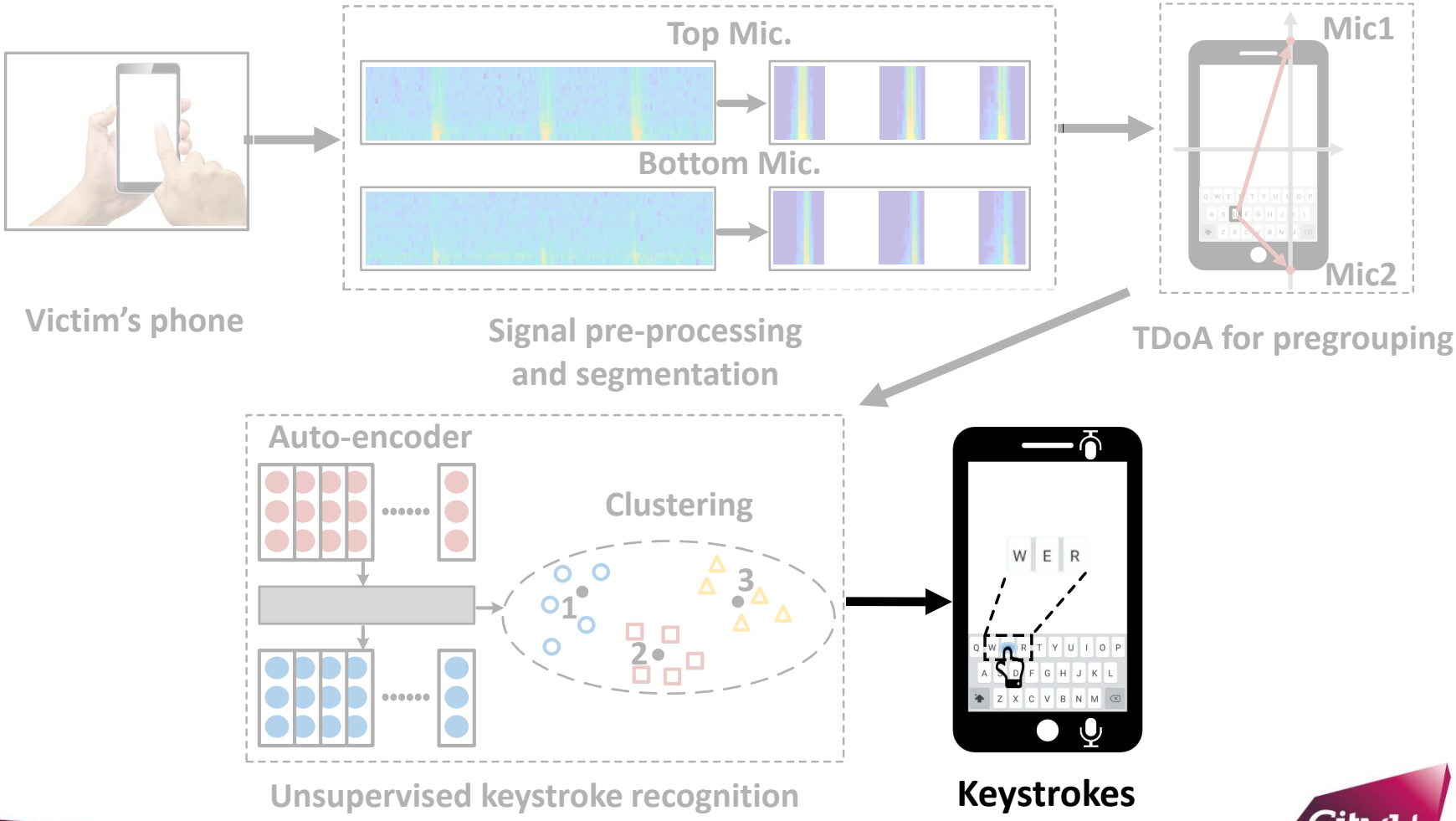


Challenge-II: unlabeled data

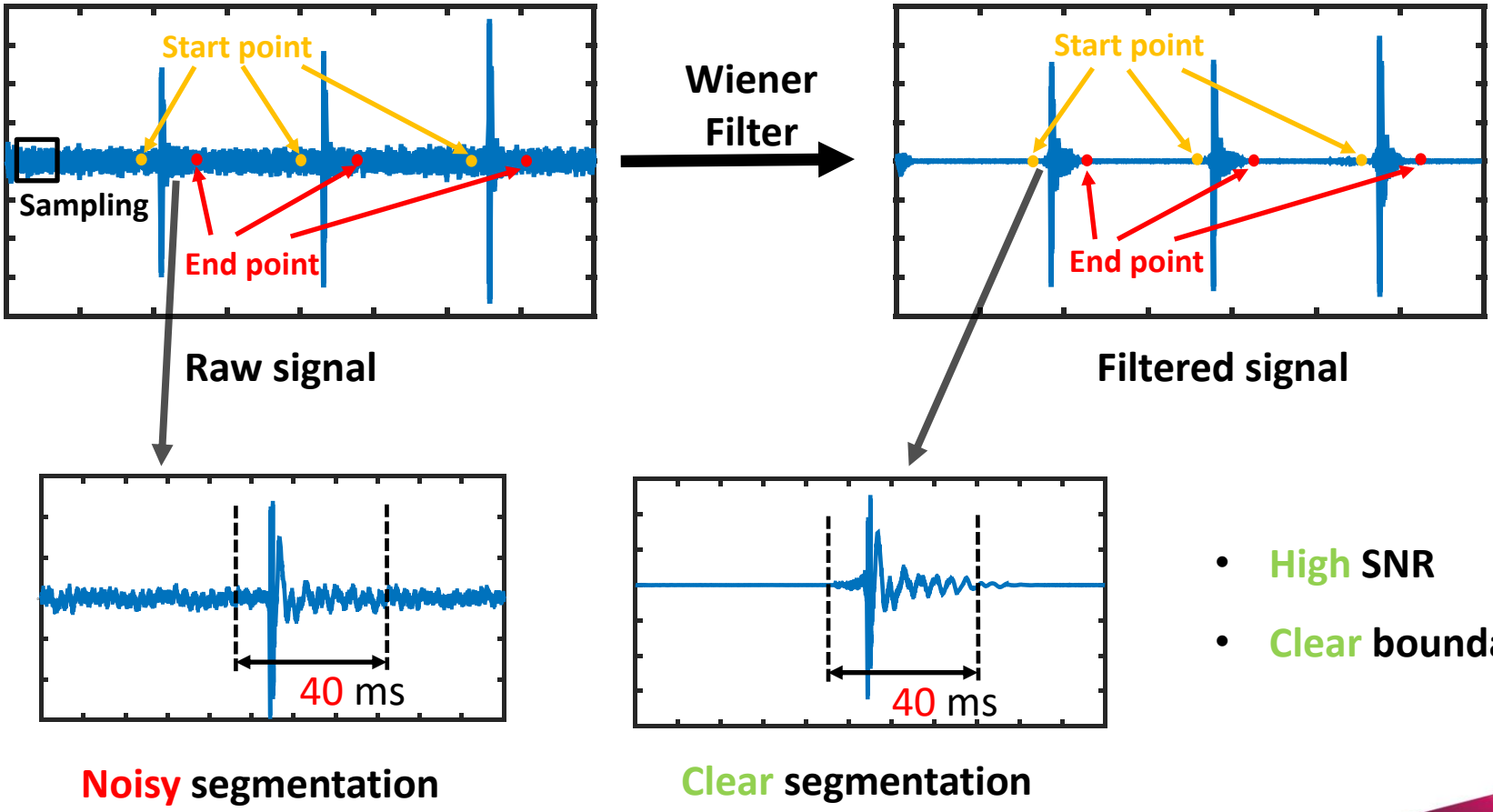


How to classify?

System overview

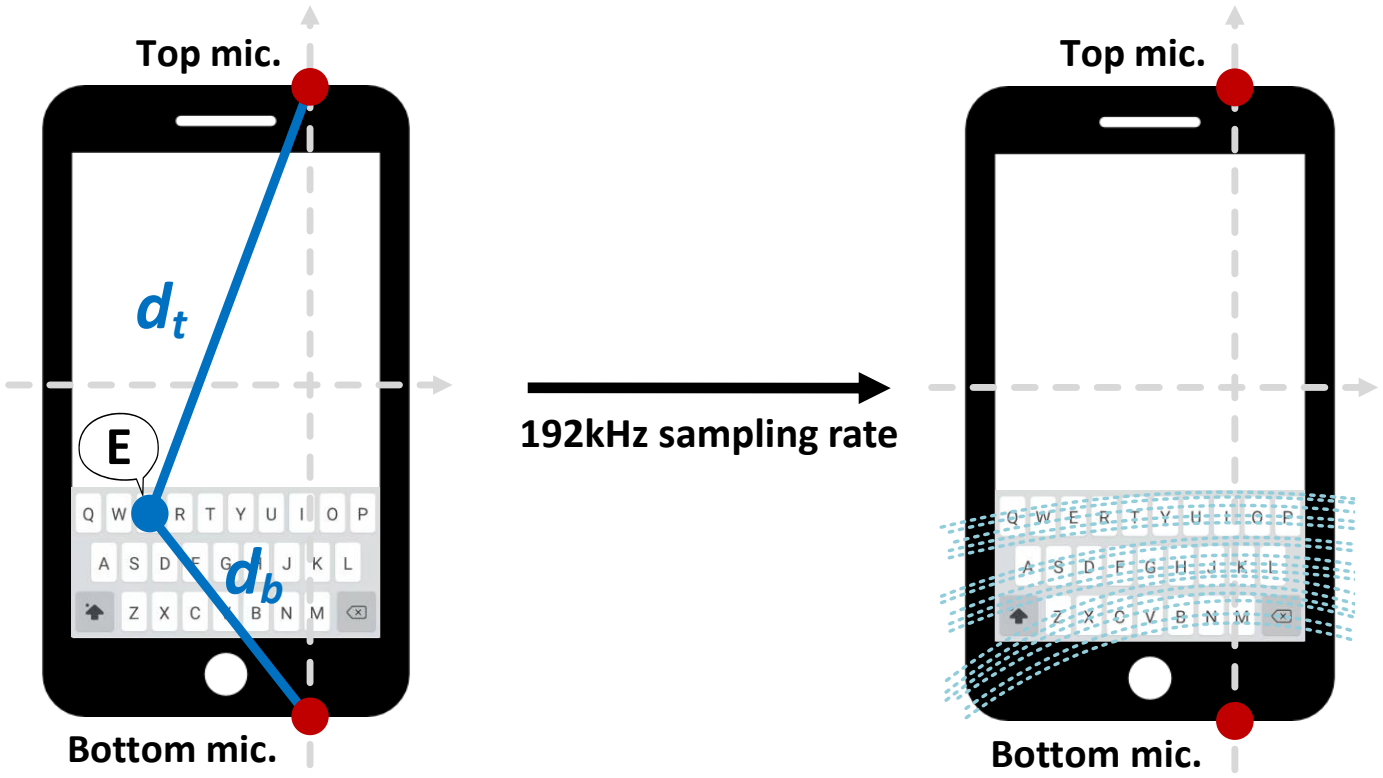


Keystroke recognition: weak signal



- High SNR
- Clear boundary

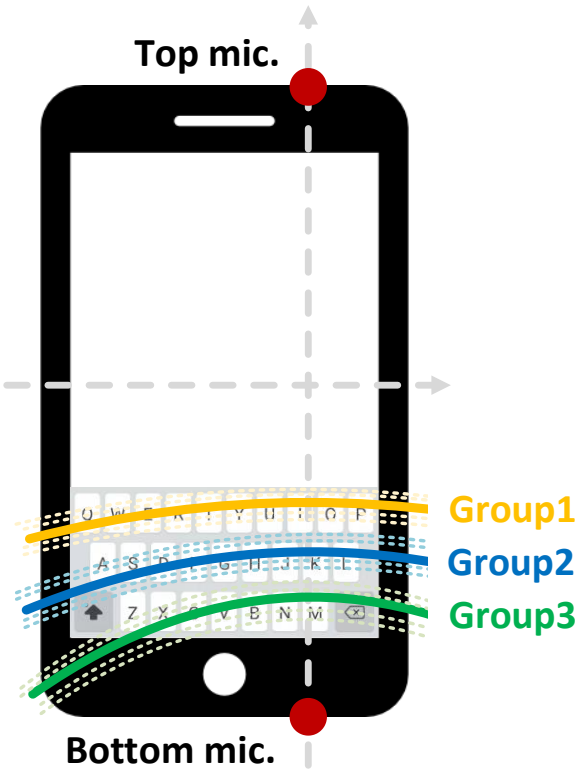
Keystroke recognition: pre-grouping

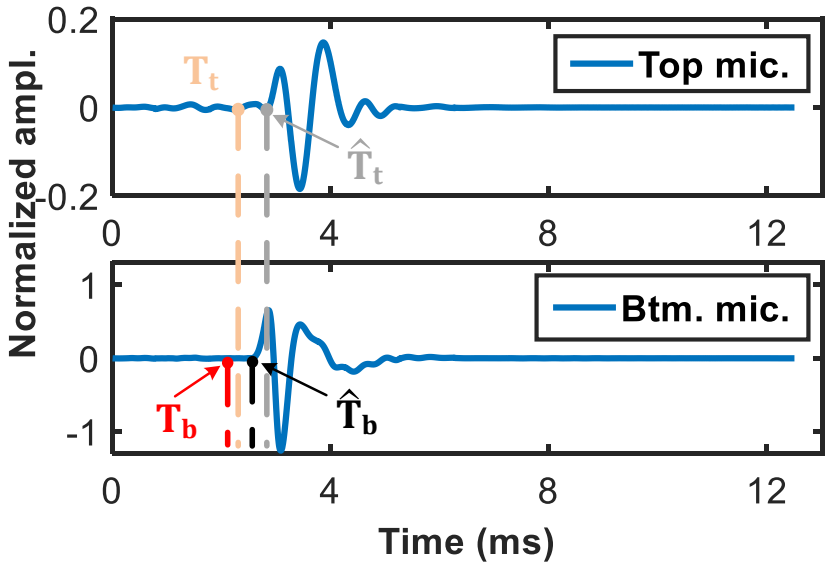


$$\Delta t = (d_t - d_b)/v$$

Connect points with the same TDoA

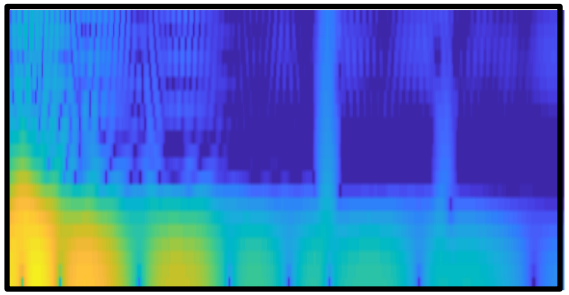
Keystroke recognition: pre-grouping



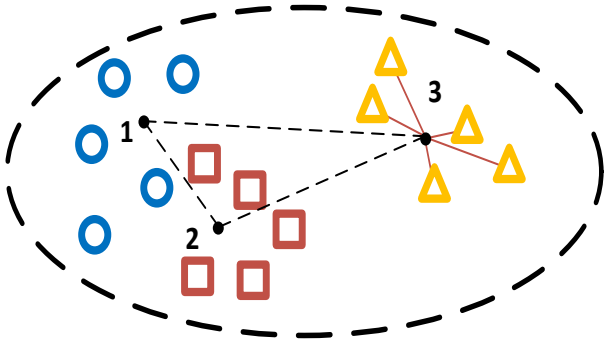


$$Error = (T_t - T_b) - (\widehat{T_t} - \widehat{T_b})$$

Unsupervised classification: clustering



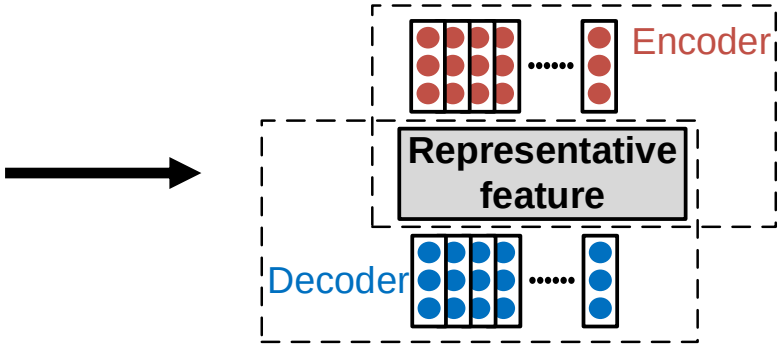
MFCC feature



Clustering

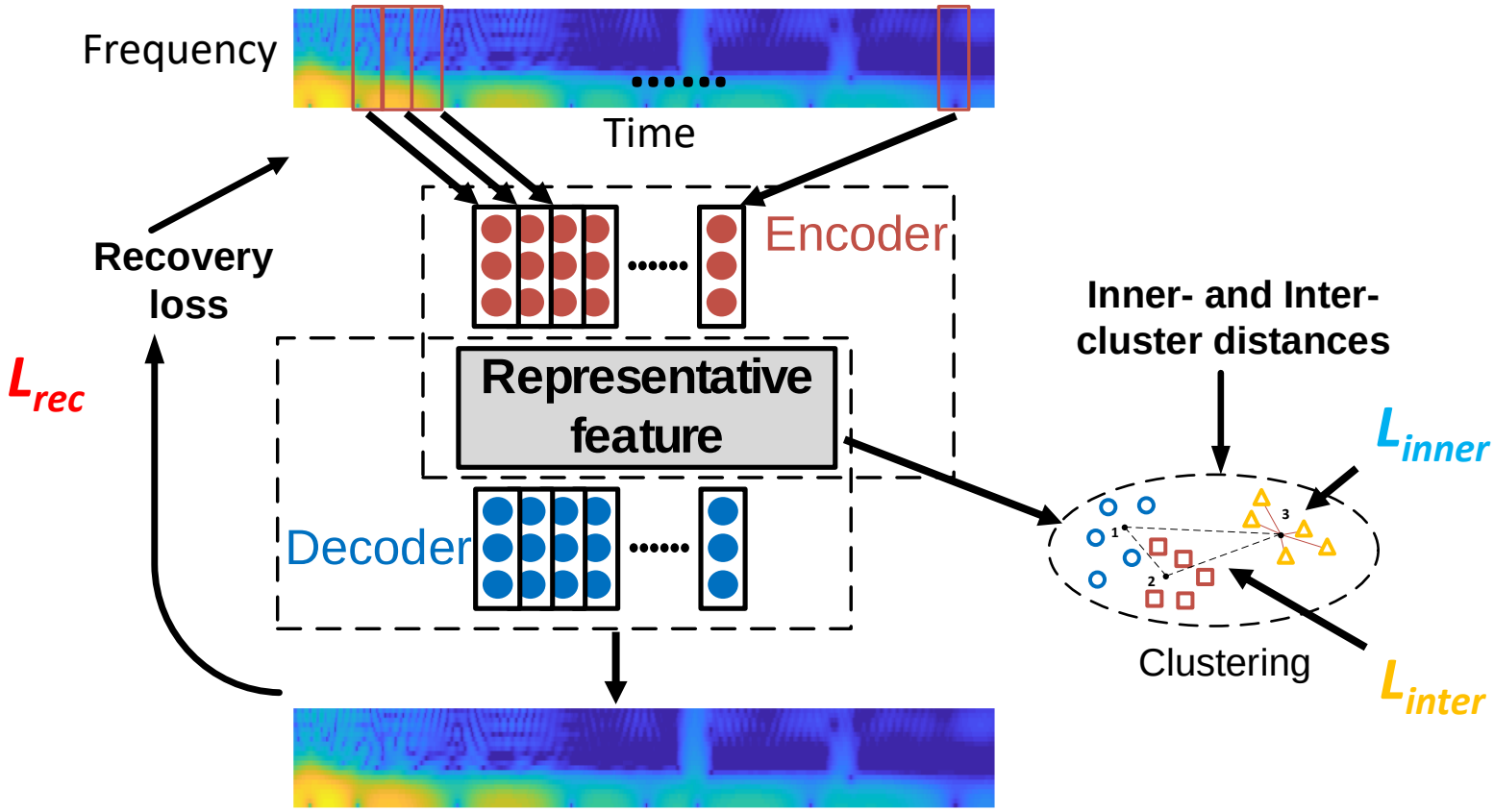
Low accuracy?

Need **better** features



Auto-encoder

Auto-encoder based clustering



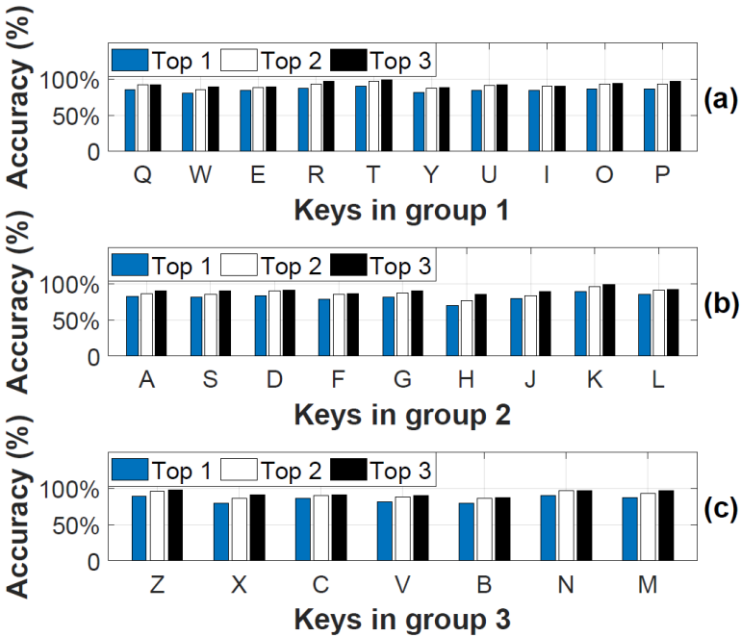
$$L_{loss} = L_{rec} + \alpha L_{inner} - \beta L_{inter}$$

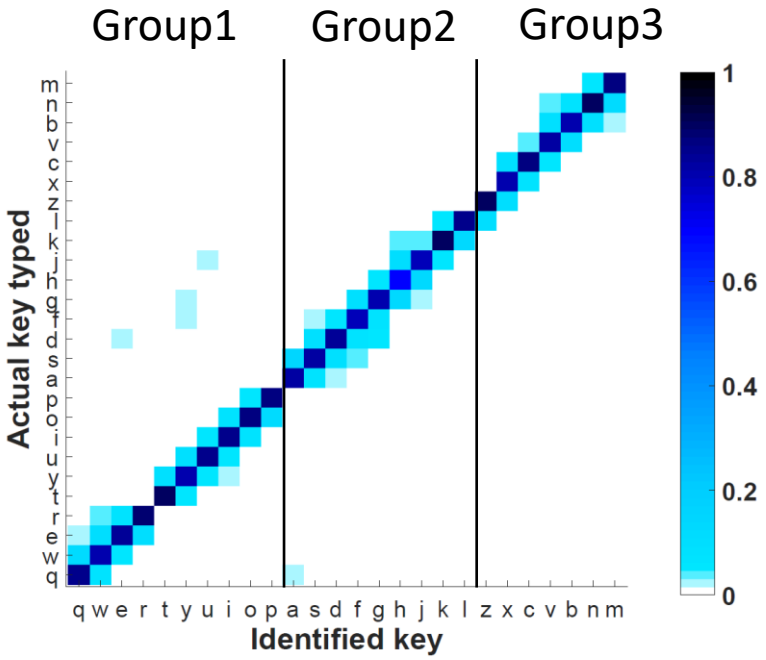
Experiment setup

- Participants: 6 volunteers as victim users
- Dataset:
  - 200 keystrokes of each key from the adversary for training
  - 4680 keystrokes from the victim users for inference
- Training: Intel i7-8700K CPU and Nvidia GTX 2080Ti GPU
- Testing: Samsung GalaxyS7, Nexus 5X and Huawei P30 Pro

Evaluation

- Overall performance

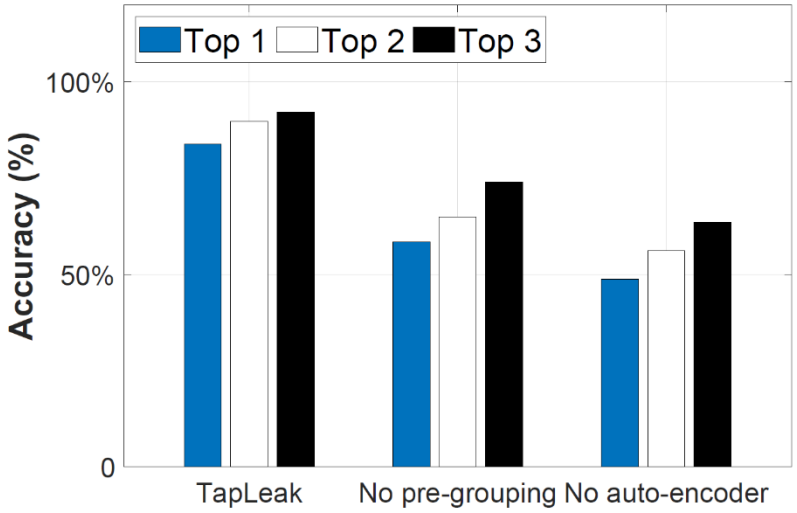




Average accuracy (top-3): 92.2%

Evaluation

- System Component

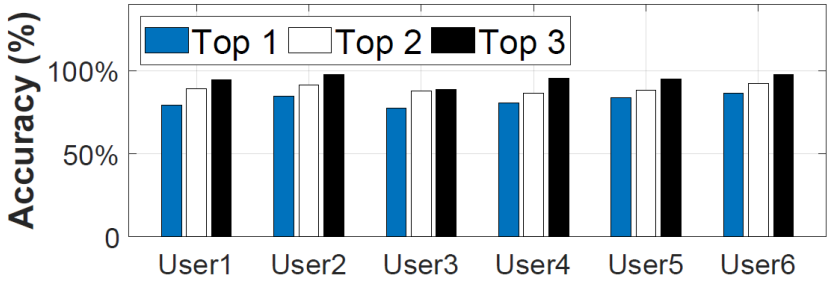


Impact of system component

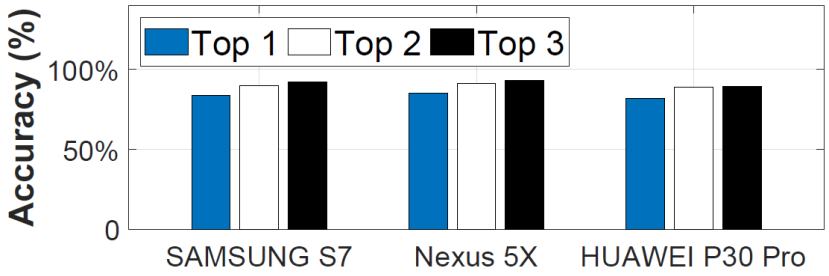
- **Accuracy: (top-3)**
  - **TapLeak: 92.9%**
  - **No pre-grouping: 74.1%**
  - **No auto-encoder: 63.5%**

Evaluation

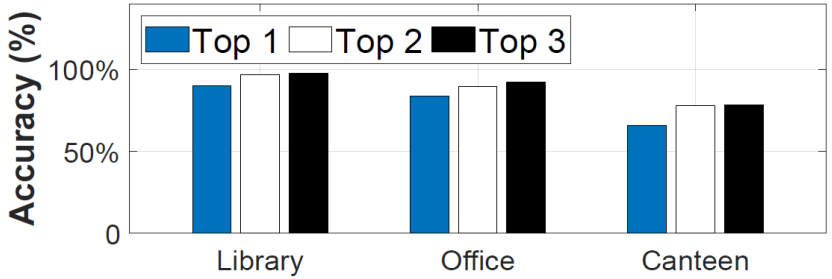
- **Different users**



- **Different phones**



- **Different noise levels**



- **Robust for different situations**

Conclusion

- Demonstrate a possible privacy leakage through dual microphones on the mobile phone
- Propose effective unsupervised model to infer keystrokes with weak acoustic signal
- Implement the prototype of the system and evaluate on different mobile phones