



Endophasia: Utilizing Acoustic-Based Imaging for Issuing Contact-Free Silent Speech Commands

Yongzhao Zhang, Wei-Hsiang Huang, Chih-Yun Yang, Wen-Ping Wang, Yi-Chao Chen^{*}, Chuang-Wen You, Da-Yuan Huang, Guangtao Xue^{*}, Jiadi Yu

^{*}Shanghai Jiao Tong University,

[^]National Taiwan University, [†]National Chiao Tung University

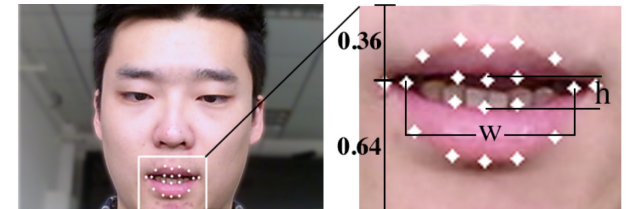
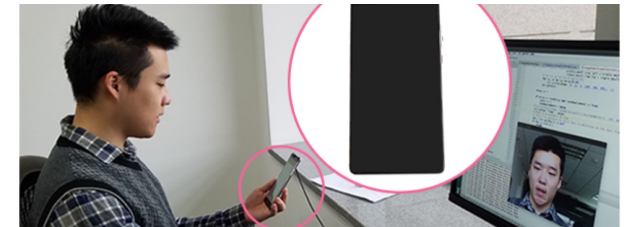
Motivation

- **Limitation of speech interface.**
 - Interference from background noise or voices from others.
 - Privacy concerns and disturbing others.
- **Target: develop a silent speech interface.**



Existing Works

- Several approaches have been proposed.
 - Invasive solution
 - No-invasive solution
 - Camera-based solution

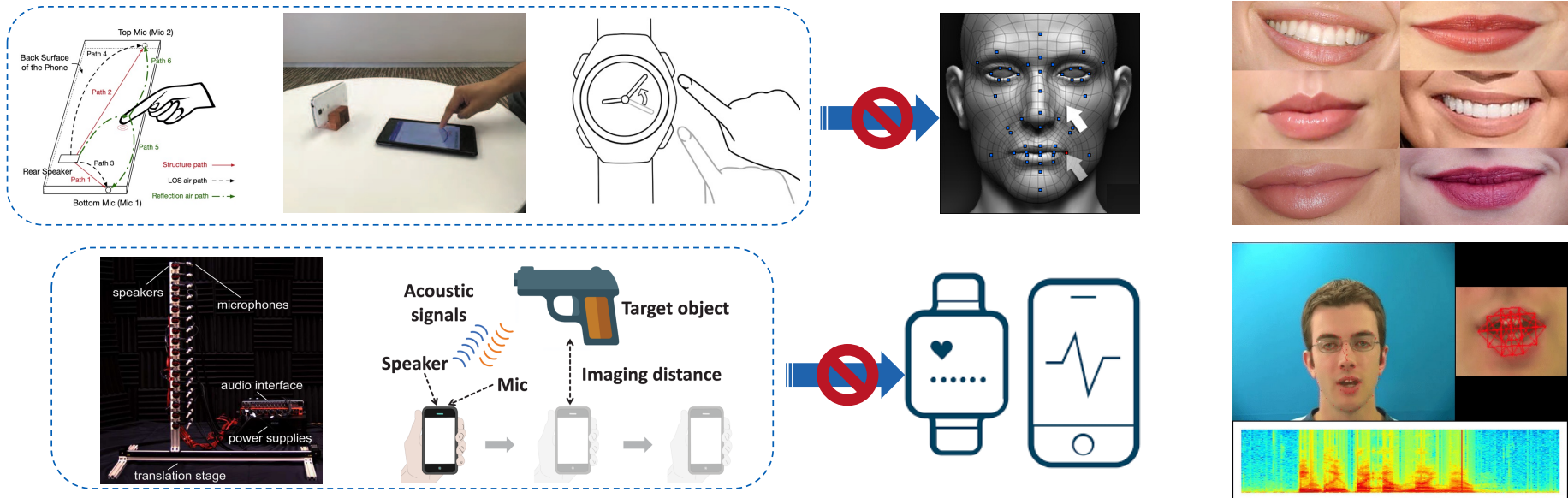


Acoustic-Based Imaging Solution!

Challenges

- **Challenges to implement acoustic-based silent speech interface.**

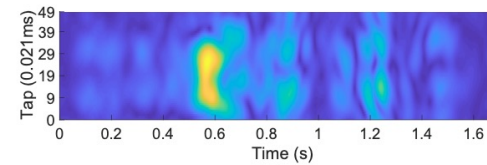
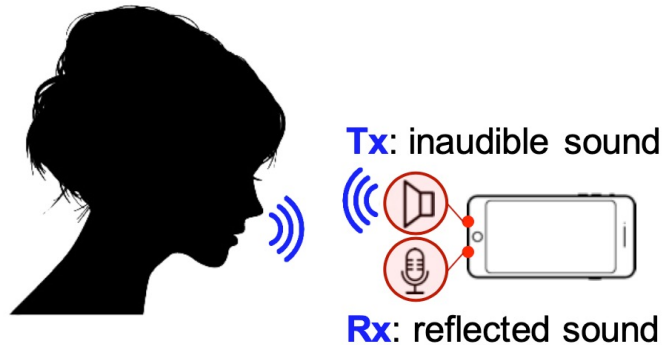
- Movement Extraction
- Limited number of mic/speaker
- User sensitivity and lack of dataset.



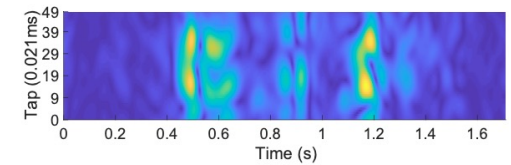
Methods

● Basic idea

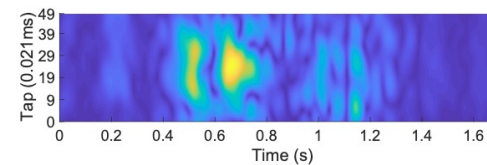
- Emit inaudible sound.
- Facial movements will change channel impulse response (CIR).
- Estimate CIR.
- Extract 2D motion profiles.
- Infer silent commands.



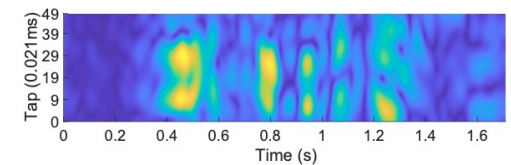
(a) Answer



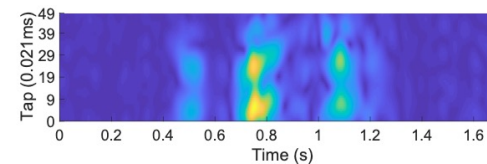
(b) Search



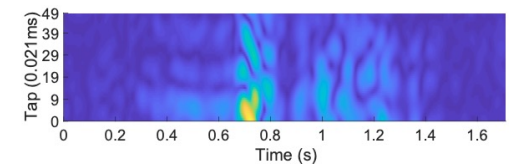
(c) Skype



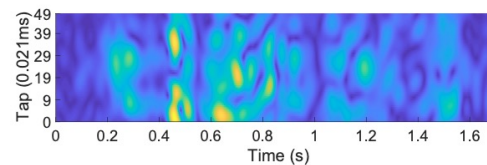
(d) Camera



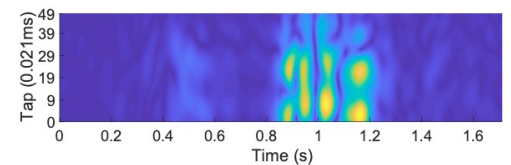
(e) Play



(f) Pause



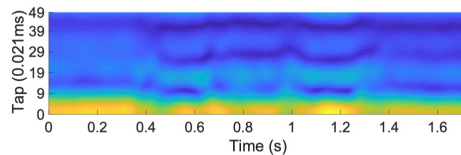
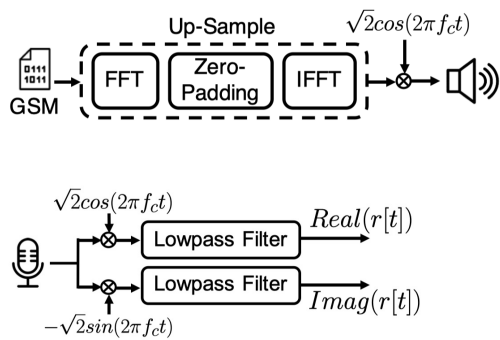
(g) Previous



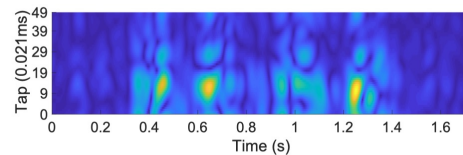
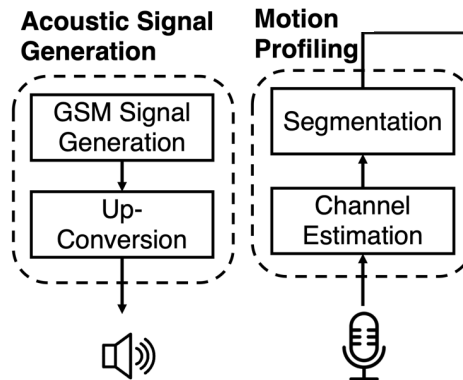
(h) Mute

Methods

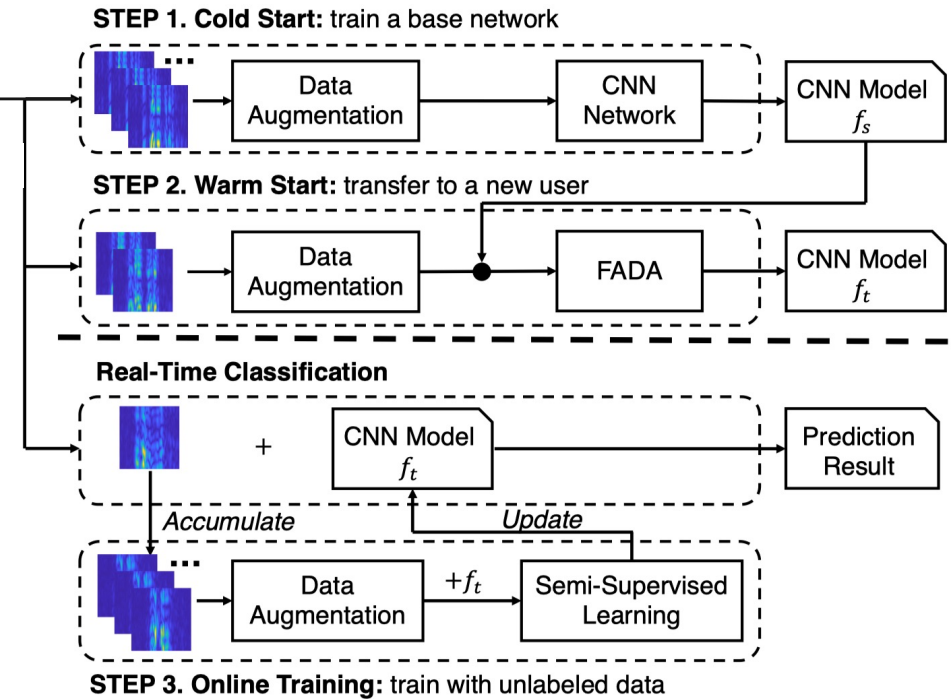
- 2D motion profiles.
- A three-step training scheme



(a) CIR.

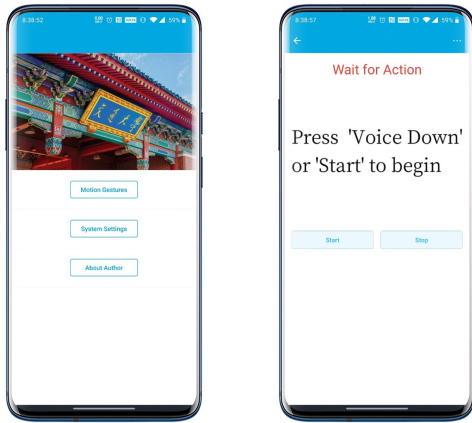


(b) Differential CIR.

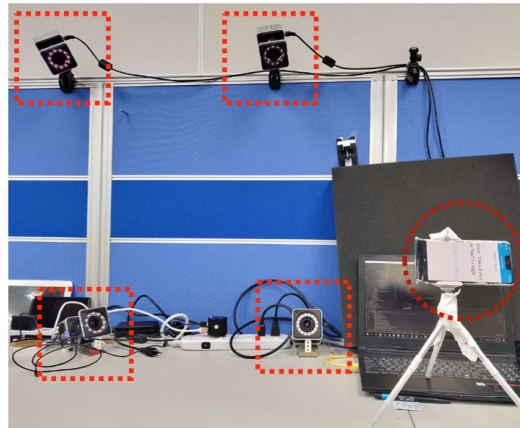


Evaluation

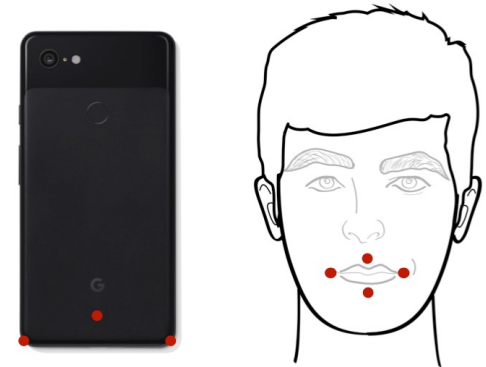
● Experiment Setup



Android App



Data Collection Configuration



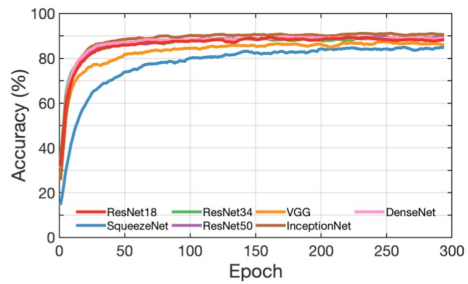
Markers

● Data Collection

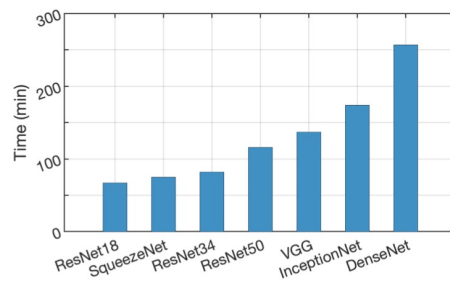
- 14 participants
- 20 commands × 30 times per each
- 17 hours of audio data

Evaluation

Micro-benchmark

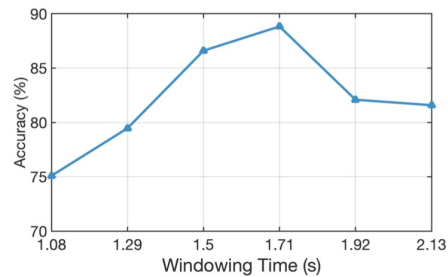
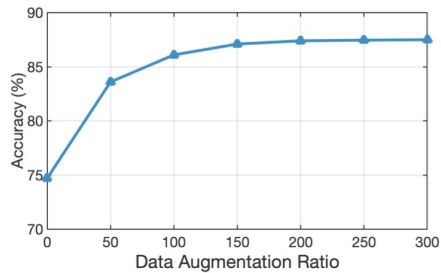


(a) Accuracy and convergence rate.



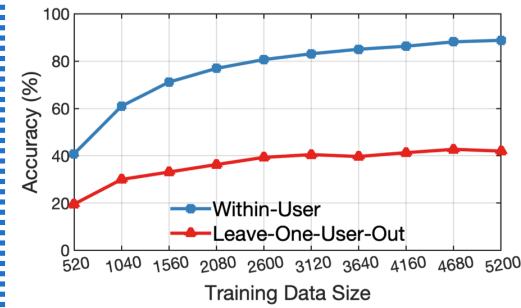
(b) Running time.

Performance of basic CNN networks

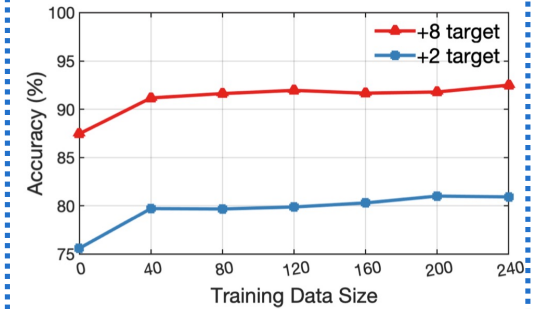


System parameters

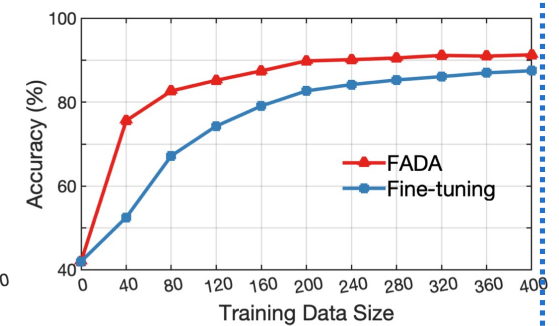
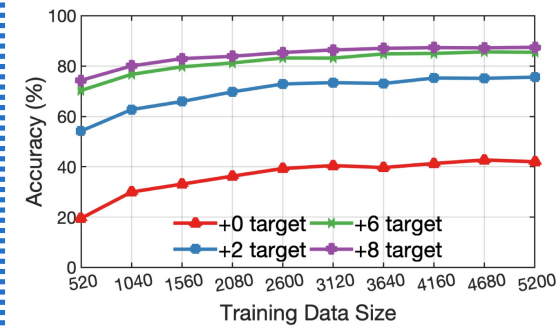
Cold Start



Online Learning



Warm Start



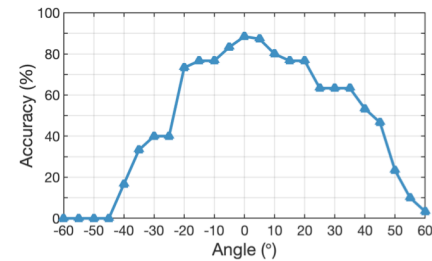
Evaluation

Confusion Matrix

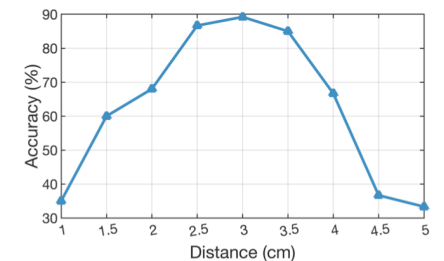
Accuracy: 92.5%

Answer-1	90.9	0.0	0.8	0.0	0.0	0.0	0.0	0.8	0.0	0.8	2.3	0.0	1.6	0.0	0.0	0.0	0.8	0.0	0.9	
Call-2	0.8	91.8	0.0	0.8	0.8	0.8	0.9	0.0	1.7	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Camera-3	0.8	0.0	93.3	0.0	0.0	0.0	1.7	1.6	0.0	0.0	0.0	0.0	0.8	0.0	0.0	2.6	0.0	0.0	0.0	
Check-4	1.7	0.8	0.0	92.5	0.0	0.8	0.9	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.9	0.0	0.0	0.9	
Copy-5	0.0	0.0	0.0	0.0	89.8	0.0	0.9	3.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	
Cut-6	0.8	0.0	0.0	0.8	1.6	91.7	0.0	0.8	0.0	0.0	1.6	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.9	
HangUp-7	1.7	0.8	0.0	0.8	2.4	0.8	91.5	1.6	0.0	0.0	0.0	0.8	0.0	0.8	0.0	0.8	0.9	0.0	0.0	
Help-8	0.0	1.6	0.0	0.0	3.1	0.8	0.0	89.3	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.8	0.8	0.9	0.0	
Home-9	0.0	2.5	0.0	0.0	0.0	0.0	1.7	0.0	97.4	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	
Mute-10	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	92.5	1.6	0.0	3.3	1.6	0.0	0.0	0.0	0.0	0.0	
Paste-11	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	91.5	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	
Pause-12	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	1.7	0.8	96.4	2.4	2.4	0.0	0.8	0.0	0.9	0.9	0.0	
Play-13	0.0	0.8	0.0	0.8	0.0	0.0	0.0	0.0	0.8	0.0	2.7	89.4	0.0	1.6	0.0	0.8	0.8	0.0	0.0	
Previous-14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	92.1	0.8	0.9	0.8	0.0	0.0	0.0	
Redial-15	0.0	0.8	0.0	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	93.5	0.0	0.8	0.0	0.0	0.0	0.0	
ScreenShot-16	0.0	0.0	2.5	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	0.0	2.4	0.8	95.7	0.0	0.0	0.0	0.0	
Search-17	1.7	0.8	0.0	0.0	0.8	0.0	0.0	0.8	0.8	0.9	0.8	0.0	0.0	0.0	92.5	0.0	0.9	0.0	0.0	
Skip-18	0.8	0.0	2.5	0.0	0.0	0.0	0.8	0.9	0.8	0.0	0.0	0.0	0.8	0.0	90.0	2.7	0.0	0.0	0.0	
Skype-19	0.0	0.0	0.0	0.0	1.6	0.8	0.9	0.0	0.0	0.8	0.0	0.0	0.8	0.0	5.8	93.8	0.0	0.0	0.0	
Undo-20	0.8	0.0	0.0	2.5	0.8	2.5	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.8	0.0	0.0	95.7	0.0	0.0	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

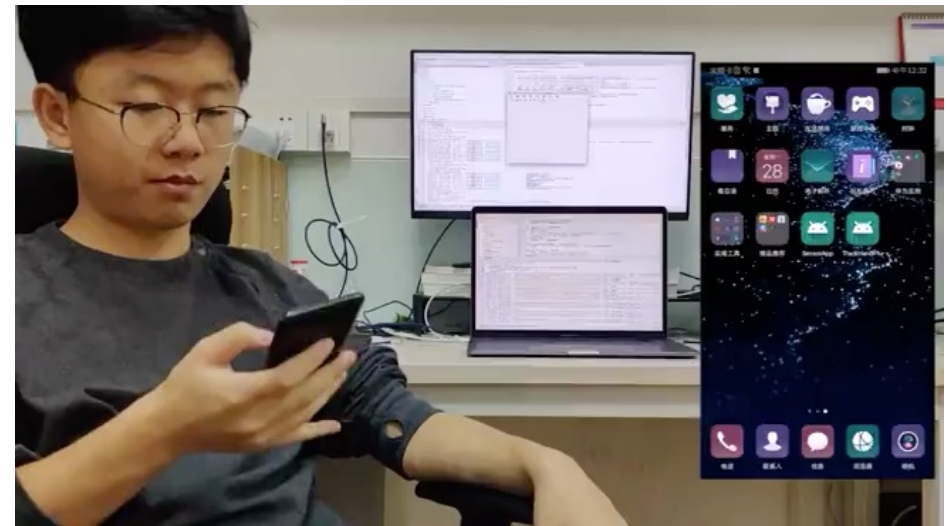
Impact of Angle



Impact of Distance



Demo Video





Thanks

