mmEve: Eavesdropping on Smartphone's Earpiece via COTS mmWave Device

Chao Wang, Feng Lin, Tiantian Liu, Kaidi Zheng, Zhibo Wang, Zhengxiong Li, Ming-Chun Huang, Wenyao Xu, Kui Ren

浙江大學 ZHEJIANG UNIVERSITY

University at Buffalo

University of Colorado Denver

昆山杜克大学 DUKE KUNSHAN UNIVERSITY

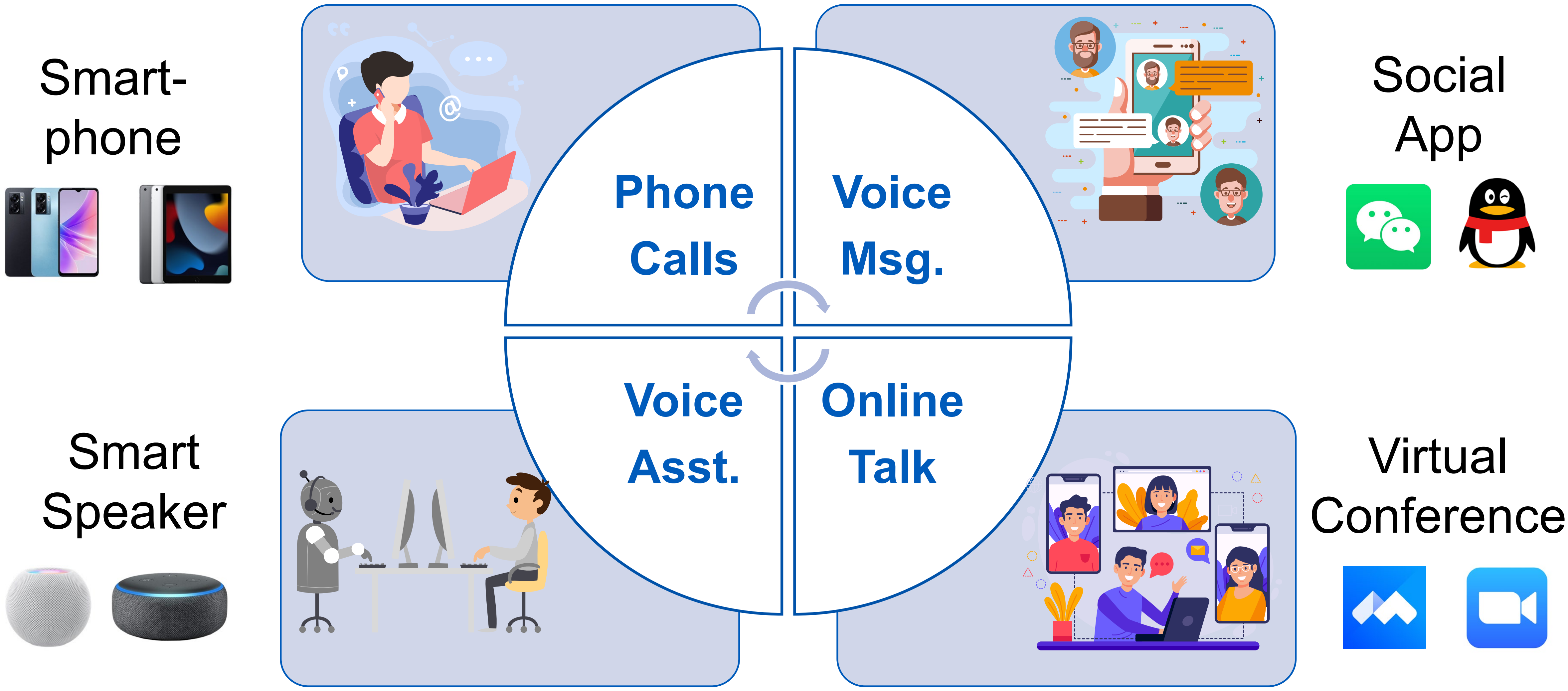
Outline

- Background
- Threat Model
- Feasibility Study
- System Design & Evaluation
- Defense
- Conclusion

Outline

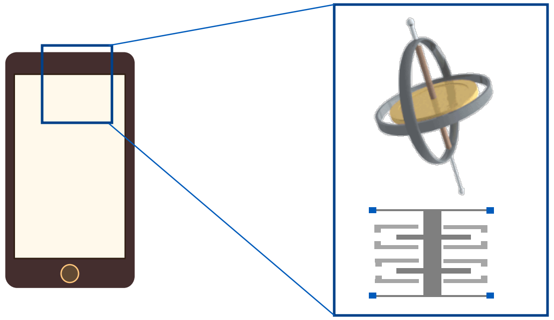
- Background
- Threat Model
- Feasibility Study
- System Design & Evaluation
- Defense
- Conclusion

Background

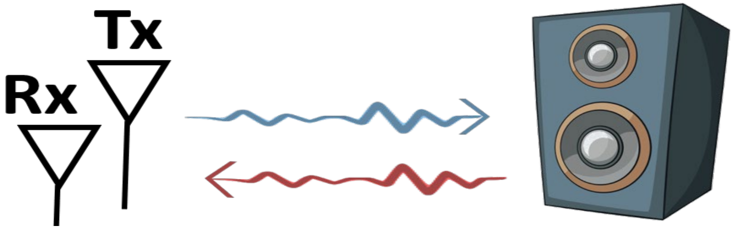

Smart-
phone
Phone
Calls
Voice
Msg.
Social
App
Smart
Speaker
Voice
Asst.
Online
Talk
Virtual
Conference

Related work

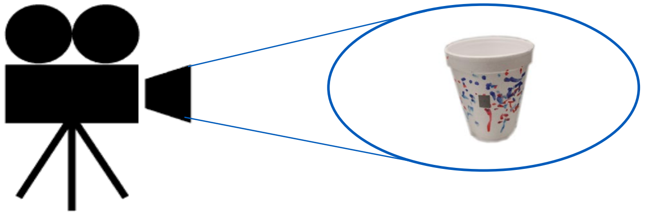
- Vibration-based eavesdropping
  - E.g., motion sensors, RF signals, video cameras, lidars…



Motion sensors (NDSS'20)



RF signals (SenSys'20)



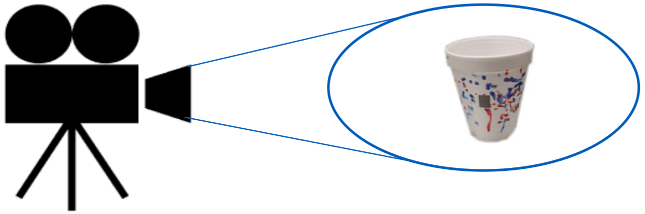
High-speed cameras (SIGGRAPH'14)

Lidar sensors (SenSys'20)

Related work

- Vibration-based eavesdropping
  - E.g., motion sensors, RF signals, video cameras, lidars…

In this work …

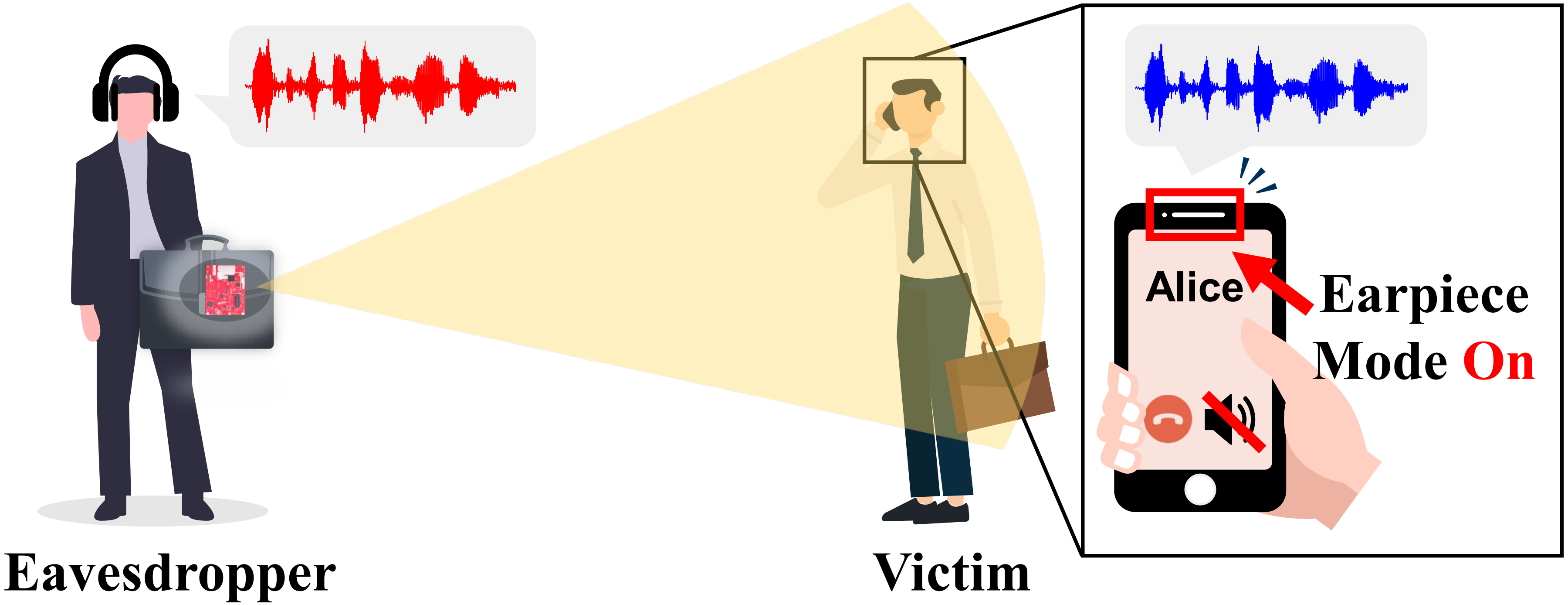


High-speed cameras (SIGGRAPH'14)

Lidar sensors (SenSys'20)

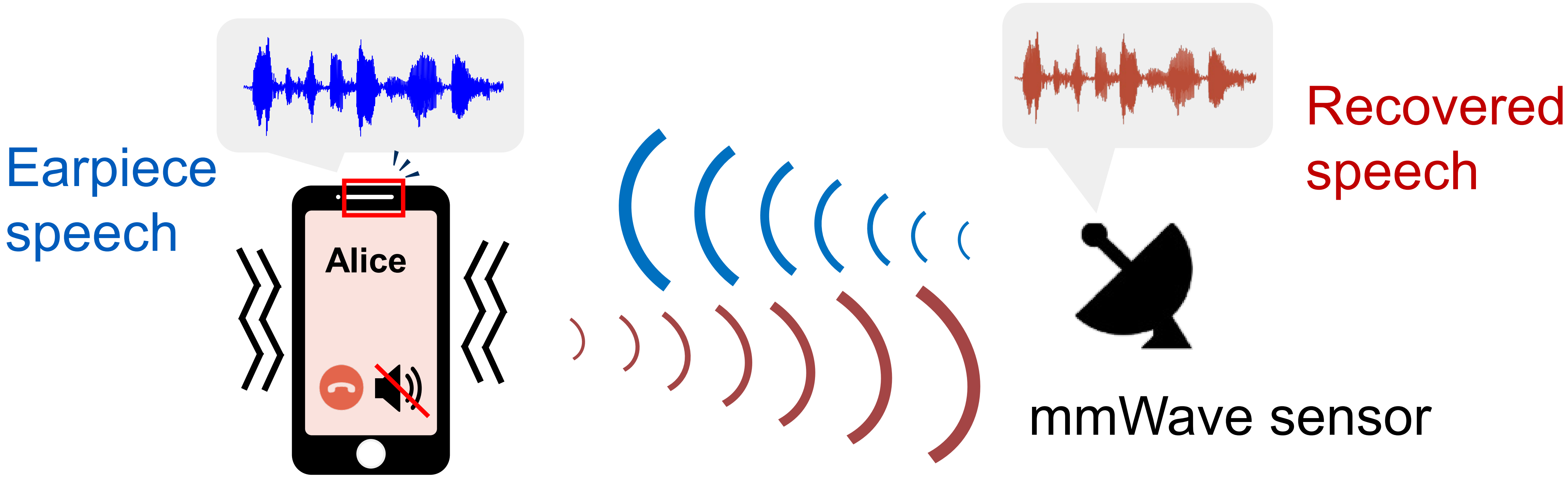
Our work

- Recover audio emitted from the **earpiece**



Principle

- **Vibration coupling** between the earpiece and the smartphone shell



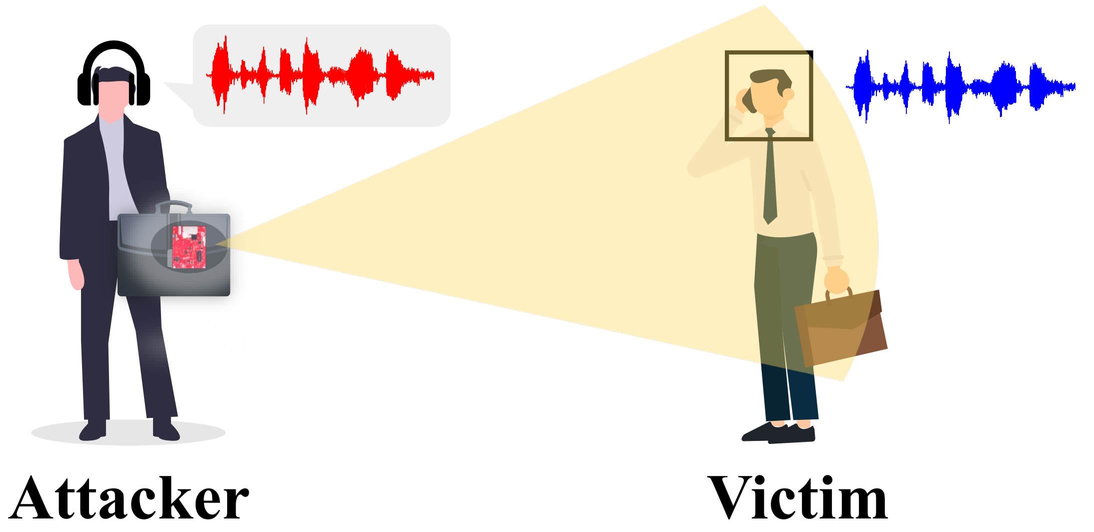
Outline

- Background
- Threat Model
- Feasibility Study
- System Design & Evaluation
- Defense
- Conclusion

Threat model

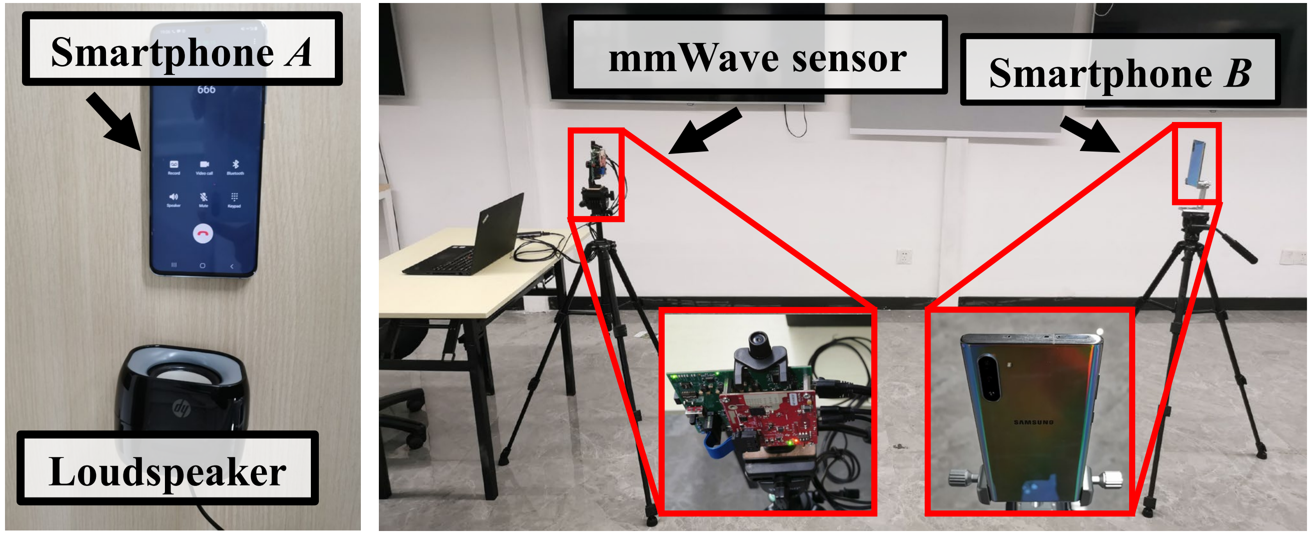
- Attack scenario
  - The victim uses the **earpiece** mode of his/her smartphone for **phone calls**/listening to **voice messages**, etc.
  - The attacker aims to recover **audible speech** of the smartphone with portable attack devices **remotely**.

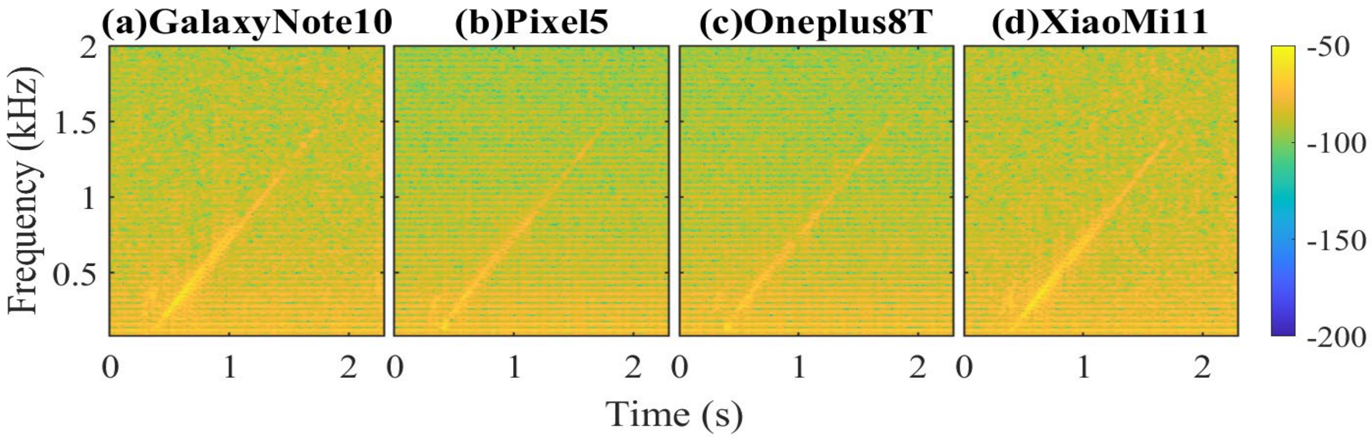
- Assumption
  - Line-of-sight condition
  - Attack distance > 2m
  - No installed malware



Feasibility study

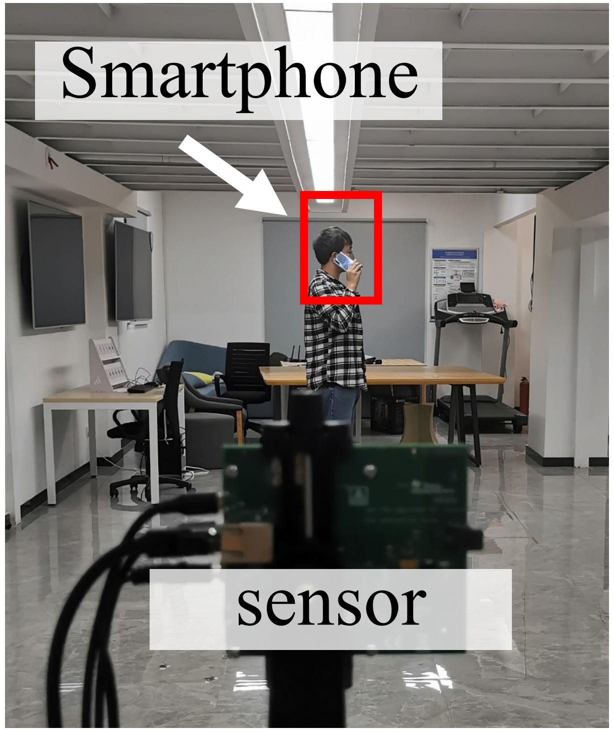
- Experimental setting
  - Phone calls
  - Audio chirp: 0-2kHz
  - Distance: 2m
- Tested smartphones
  - Galaxy Note10
  - Pixel 5
  - Oneplus 8T
  - Xiaomi 11

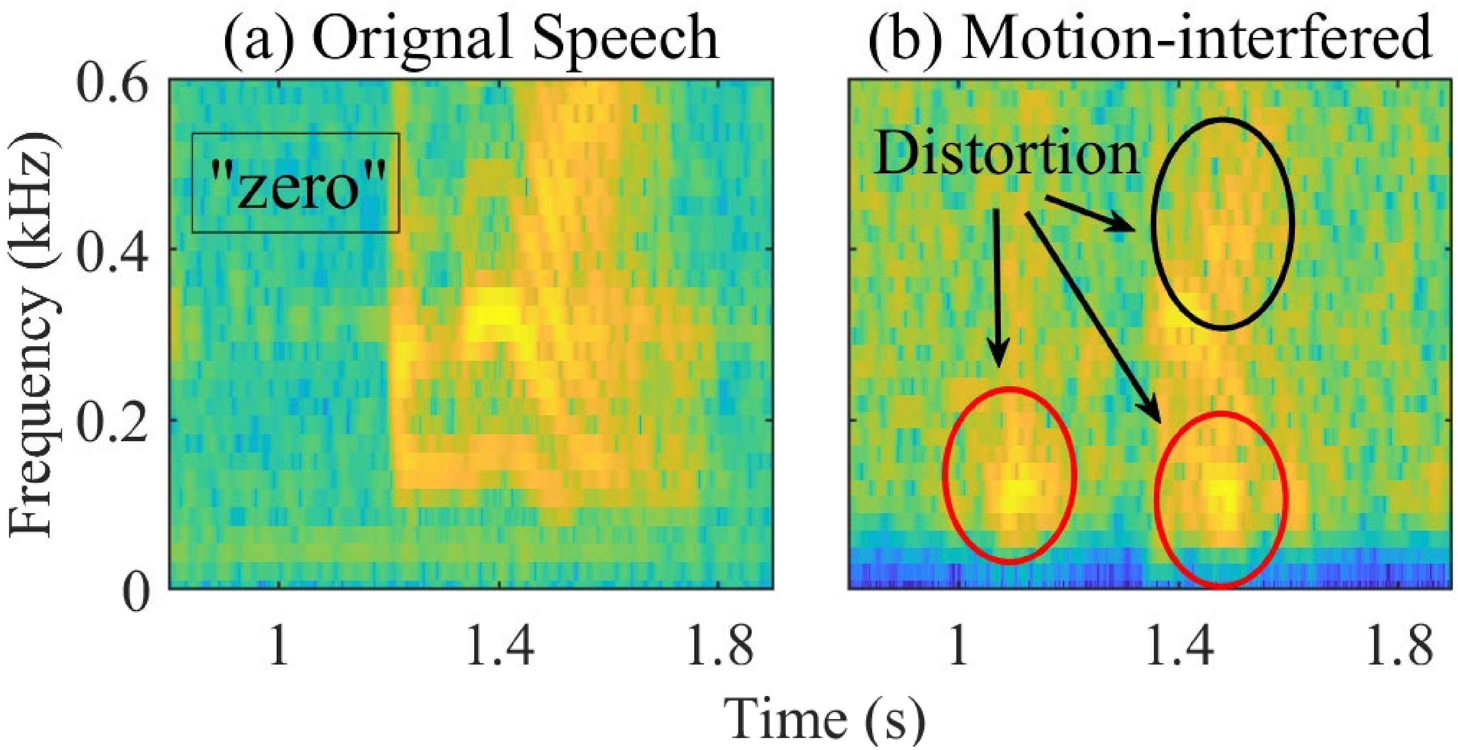




Handhold condition

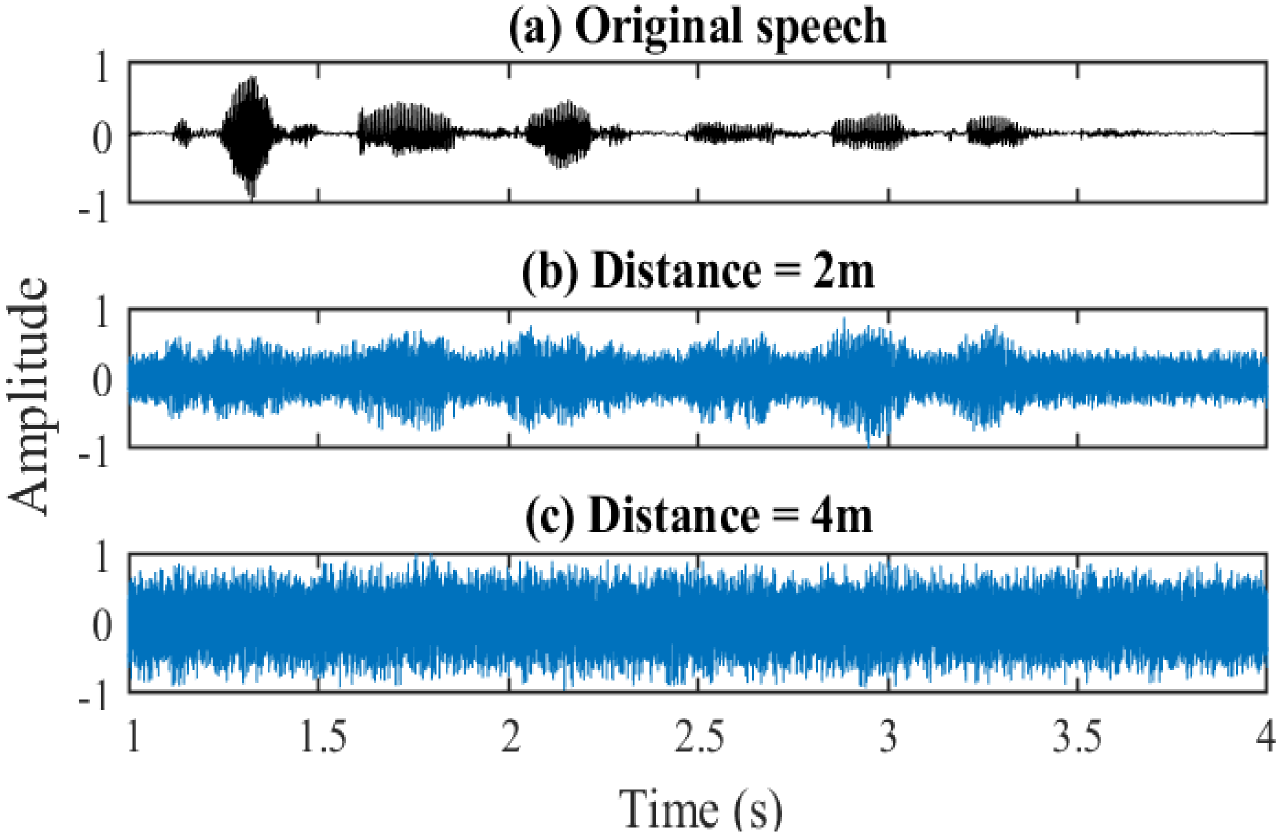
- Body movements can cause distortion on the recovered speech spectrogram.





Long-range attack

- The SNR of recovered speech signal deteriorates with the increasing sensing distance.



Tx/Rx gain

$$SNR = \frac{\alpha\lambda^2 \boxed{G_{Tx}G_{Rx}}}{(4\pi)^3 \boxed{d}^4 \boxed{F}}$$

Noise floor

Sensing distance

Summary of challenges

- **Motion interference**
- **Low SNR**

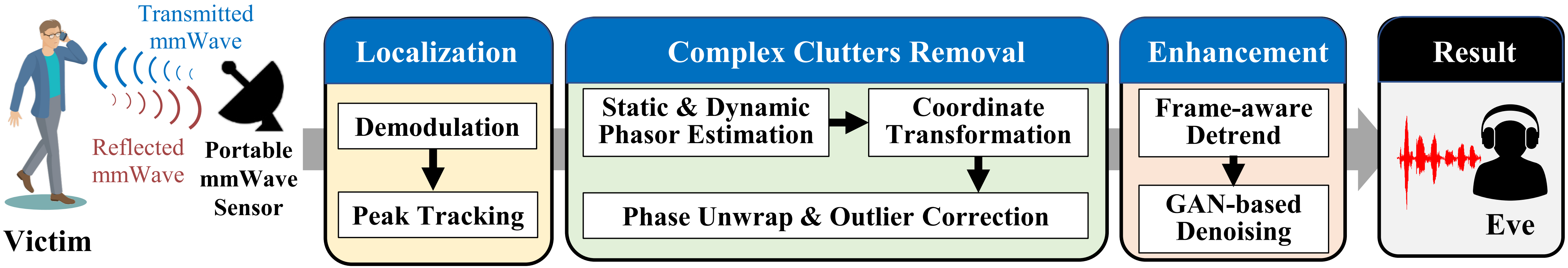
How to build a **motion-resilient** and **long-range** attack?

Outline

- Background
- Threat Model
- Feasibility Study
- System Design & Evaluation
- Defense
- Conclusion

System design

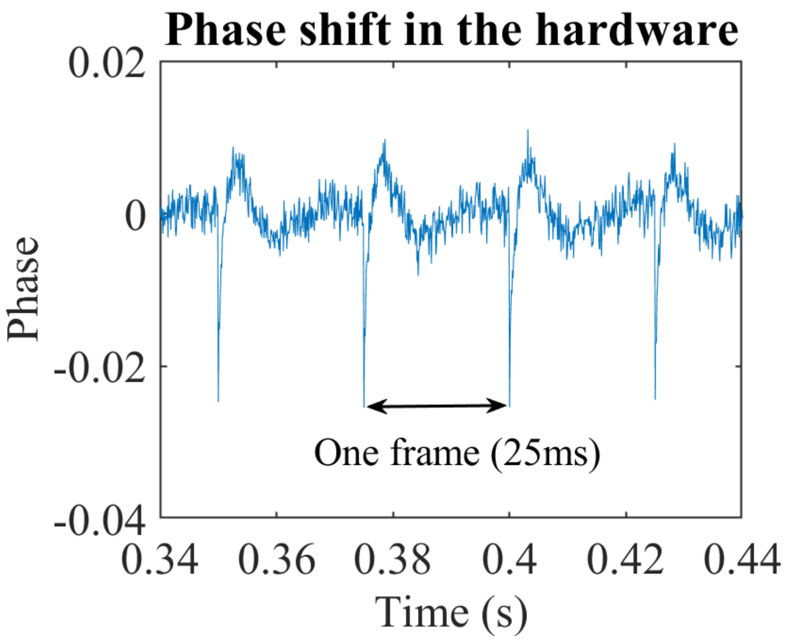
- Target localization (Range-FFT, Doppler-FFT, Angle-FFT)
- Clutter suppression (remove static/dynamic clutters)
- Speech enhancement (improve speech quality)

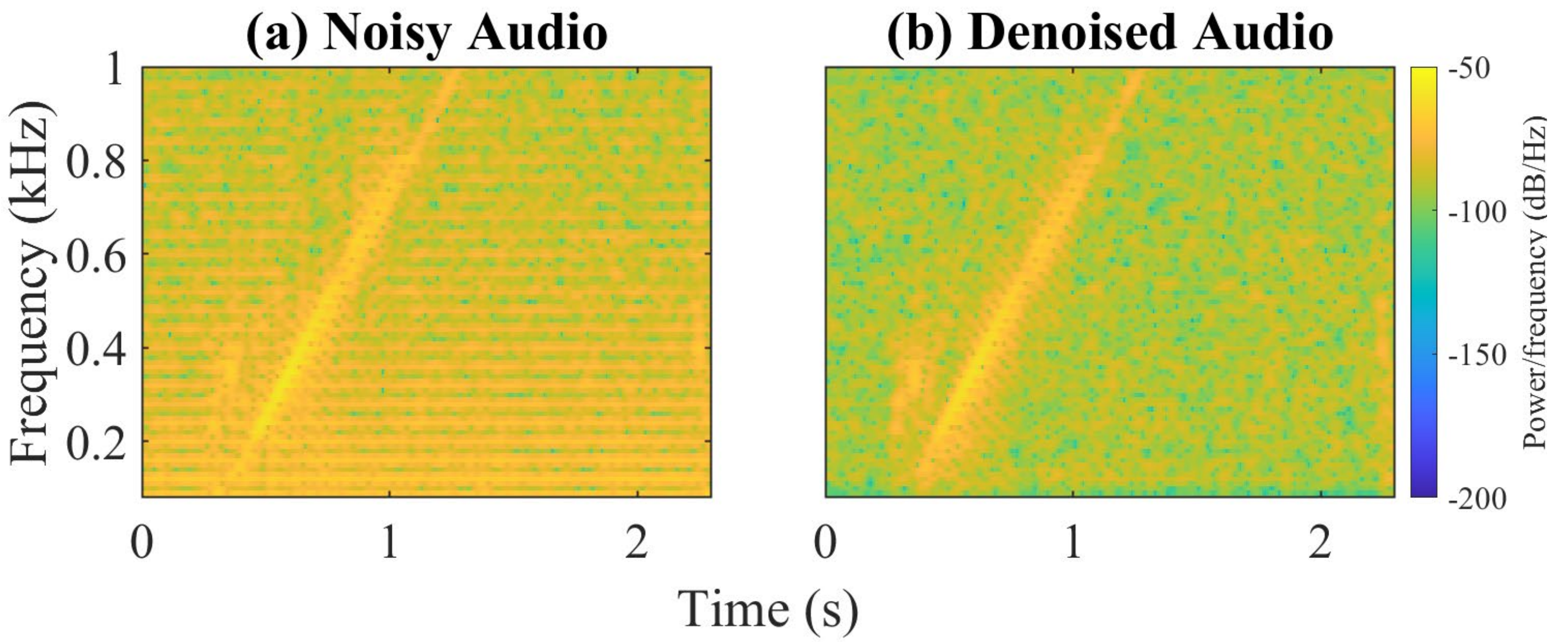


System overview

Preprocessing

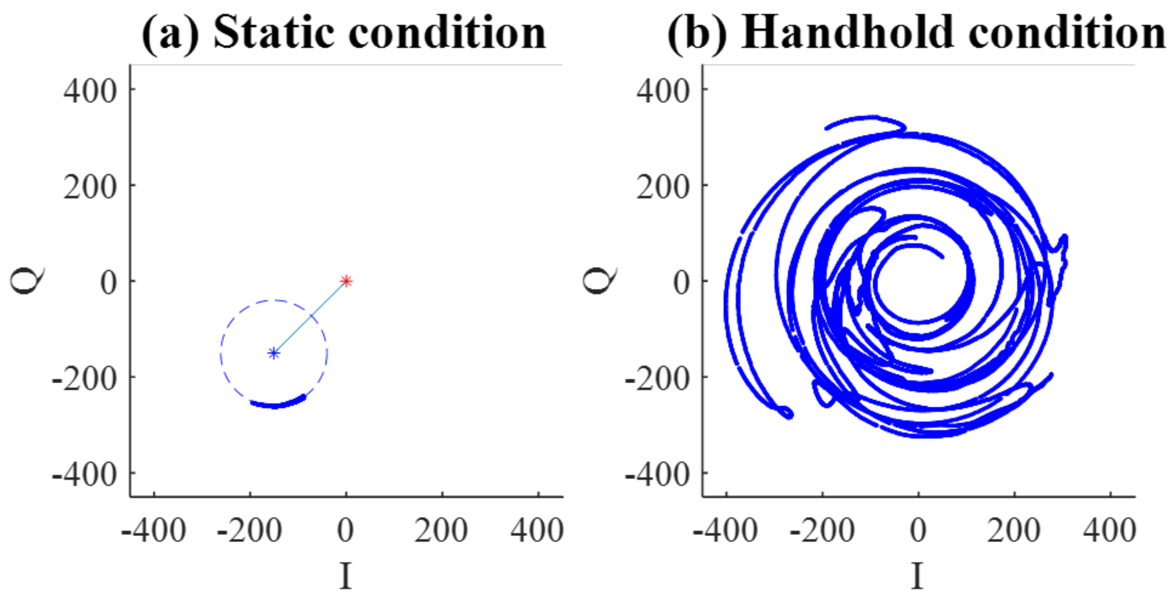
- Cause: discontinuous phases between every two frames of demodulated mmWave signals
- Solution: Frame-aware detrend $p(x) = p_1 x^n + p_2 x^{n-1} + \boxed{\ldots + p_n x + p_{n+1}}$

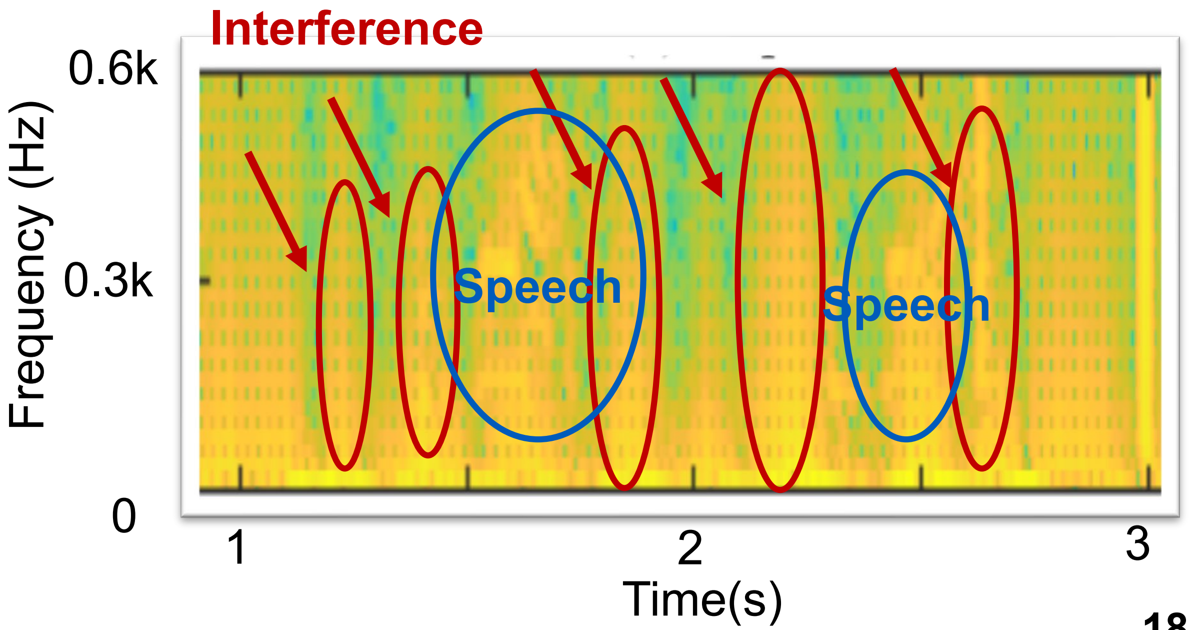




Clutter suppression

- Irregular helical curves on the I/Q plane due to human movements
- Random noise on the recovered speech spectrogram

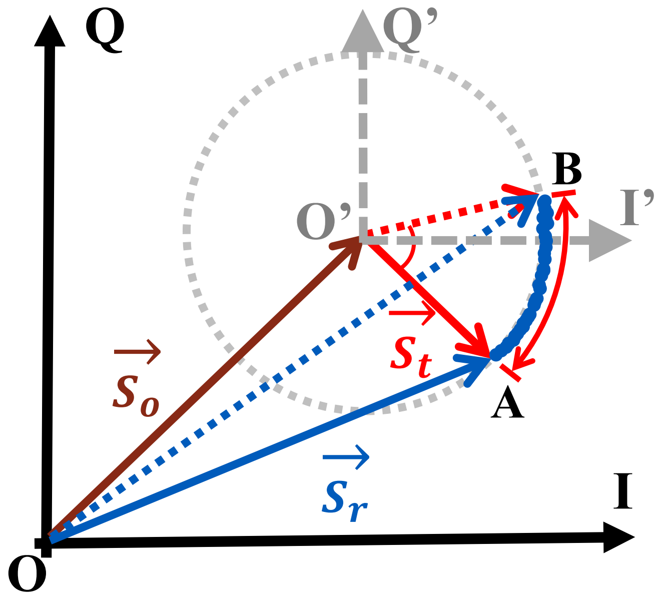


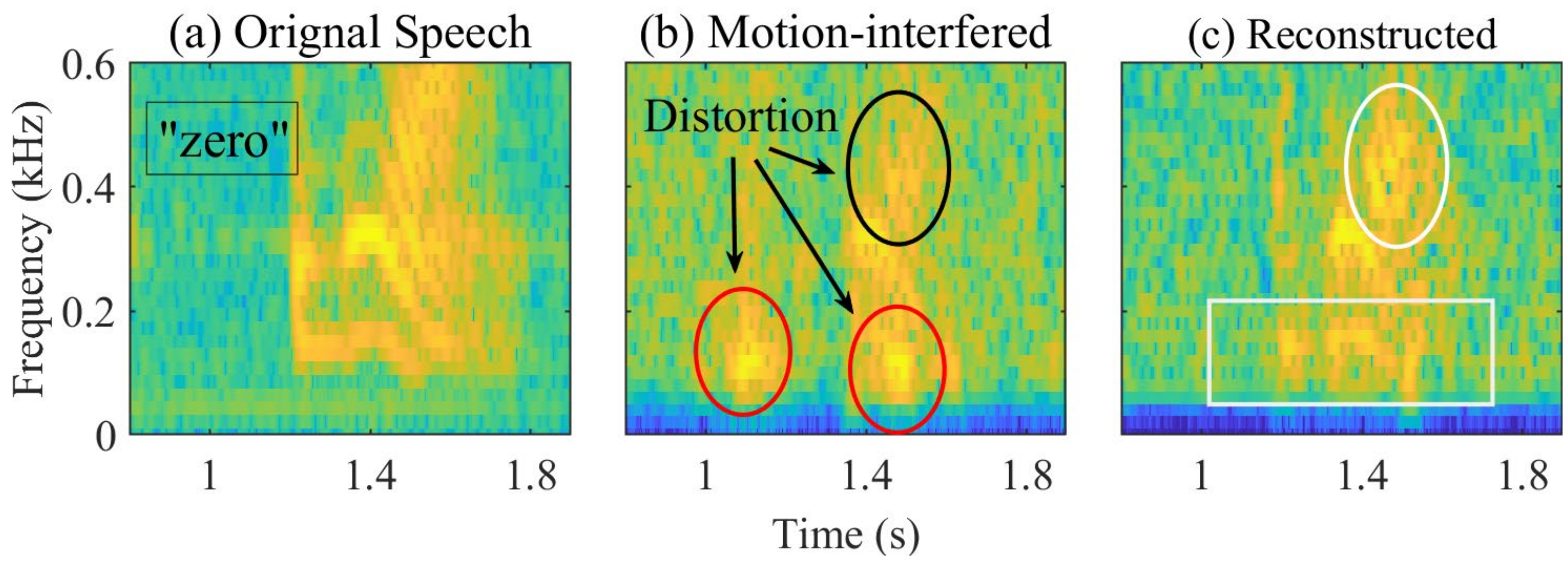


Clutter suppression

- Solution

Trajectory segmentation → Coordinate transformation → Outlier detection & rectification

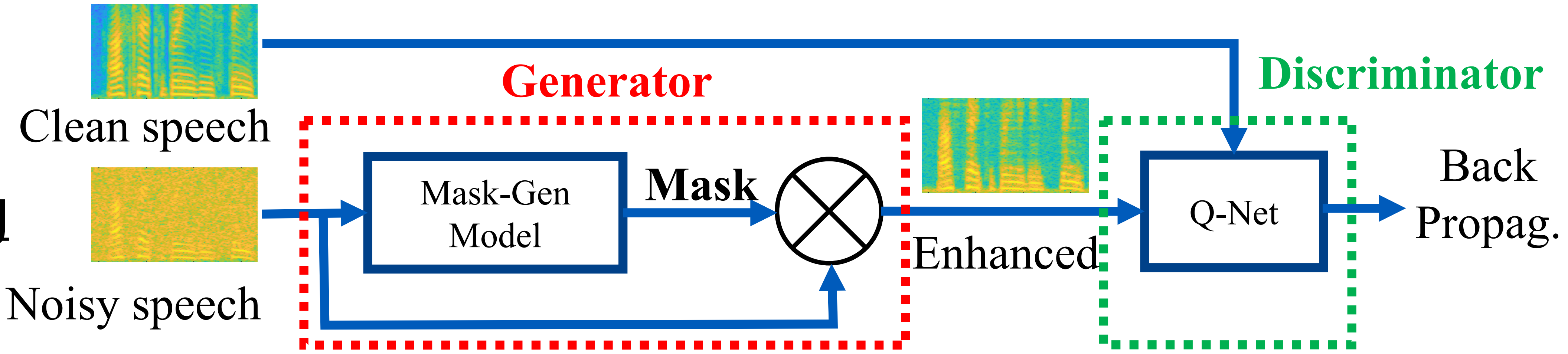




Speech enhancement

- Generative adversarial network for denoising
- Data synthesization: public audio + mmWave noise
- Enhancement: the trained Generator

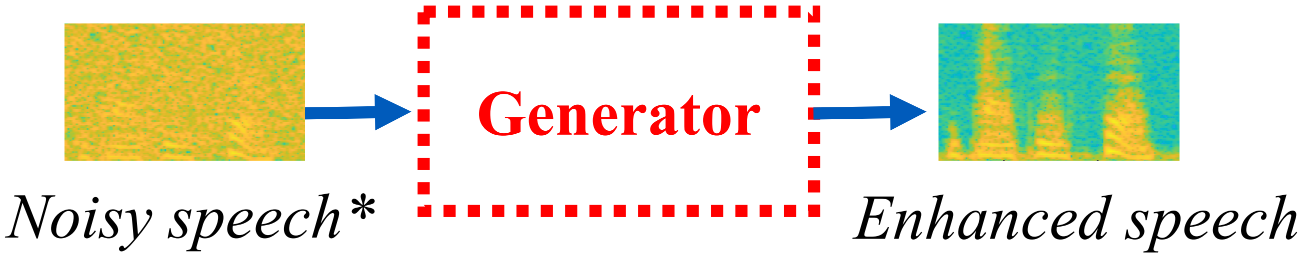
Model training



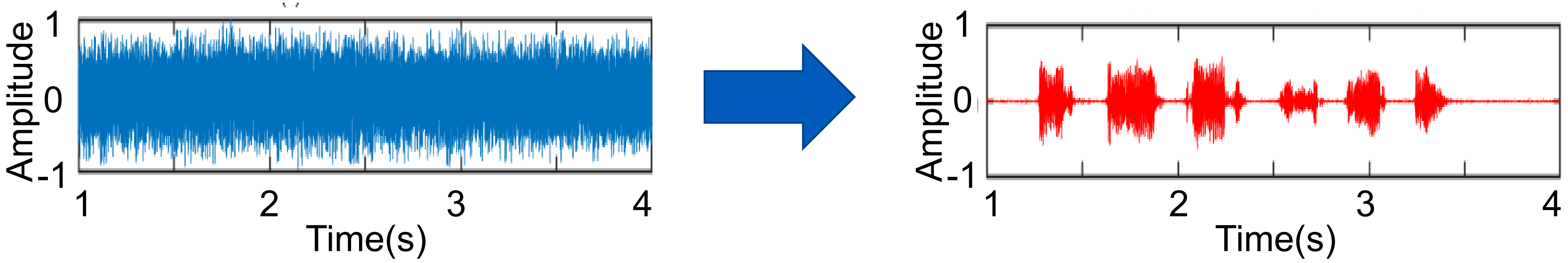
Speech enhancement

- Input: raw recovered speech after clutter suppression
- Output: the enhanced speech

Denoising phase

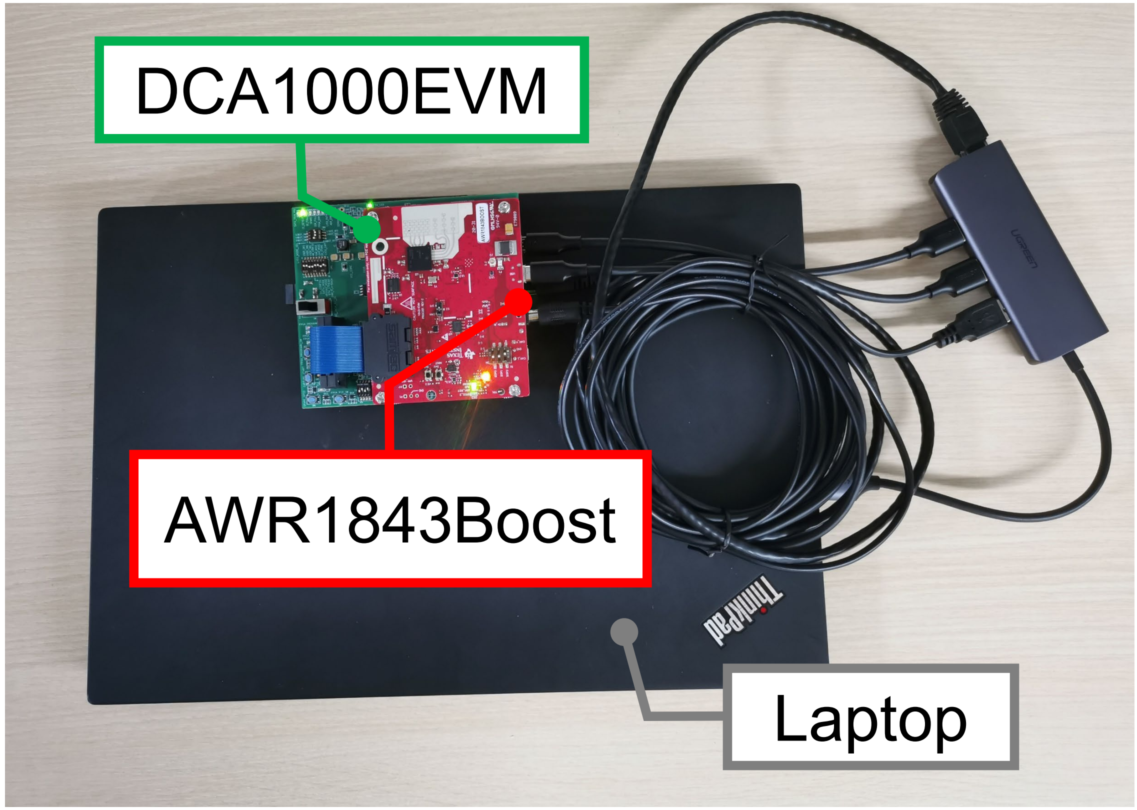


Raw recovered speech without enhancement



System setup

- Data collection
  - AWR1843Boost
  - DCA1000EVM
- Signal processing
  - Laptop（Thinkpad T490）
- Model training
  - Linux server
  - GeForce RTX 3090*4



Metrics & Dataset

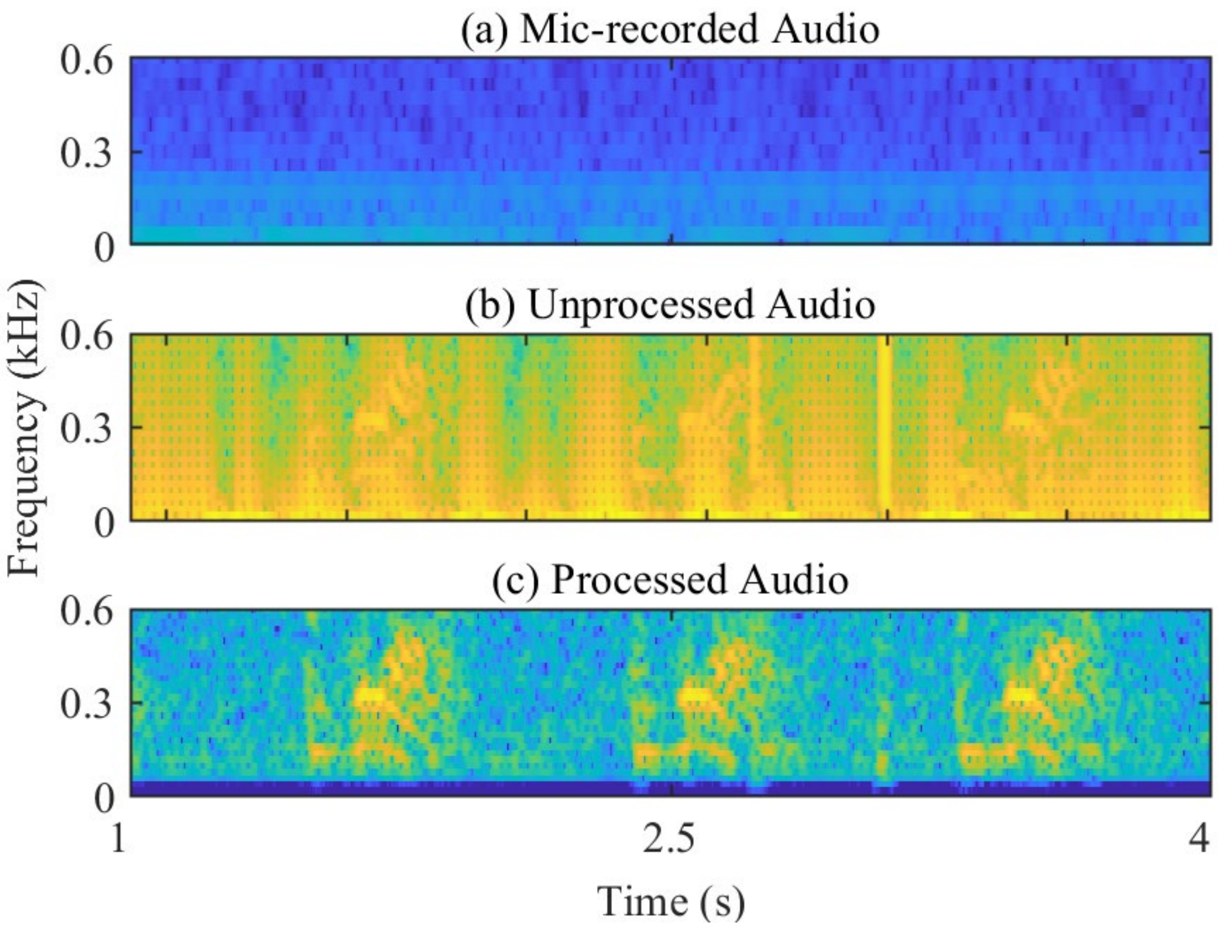
- Metric
  - Peak Signal-to-Noise Ratio (PSNR) : quantify the speech quality (a higher PSNR indicates a better speech quality)
  - Short-time Objective Intelligibility (STOI): quantify the speech intelligibility, with the score within [0,1] (the higher, the better)

- Dataset
  - Speech corpus: Harvard Sentence * 100
  - Collected from 23 different smartphone models

Sound recovery

- Recovered audio
  - Microphone (GM-S801)
  - Unprocessed (mmEve)
  - Processed (mmEve)

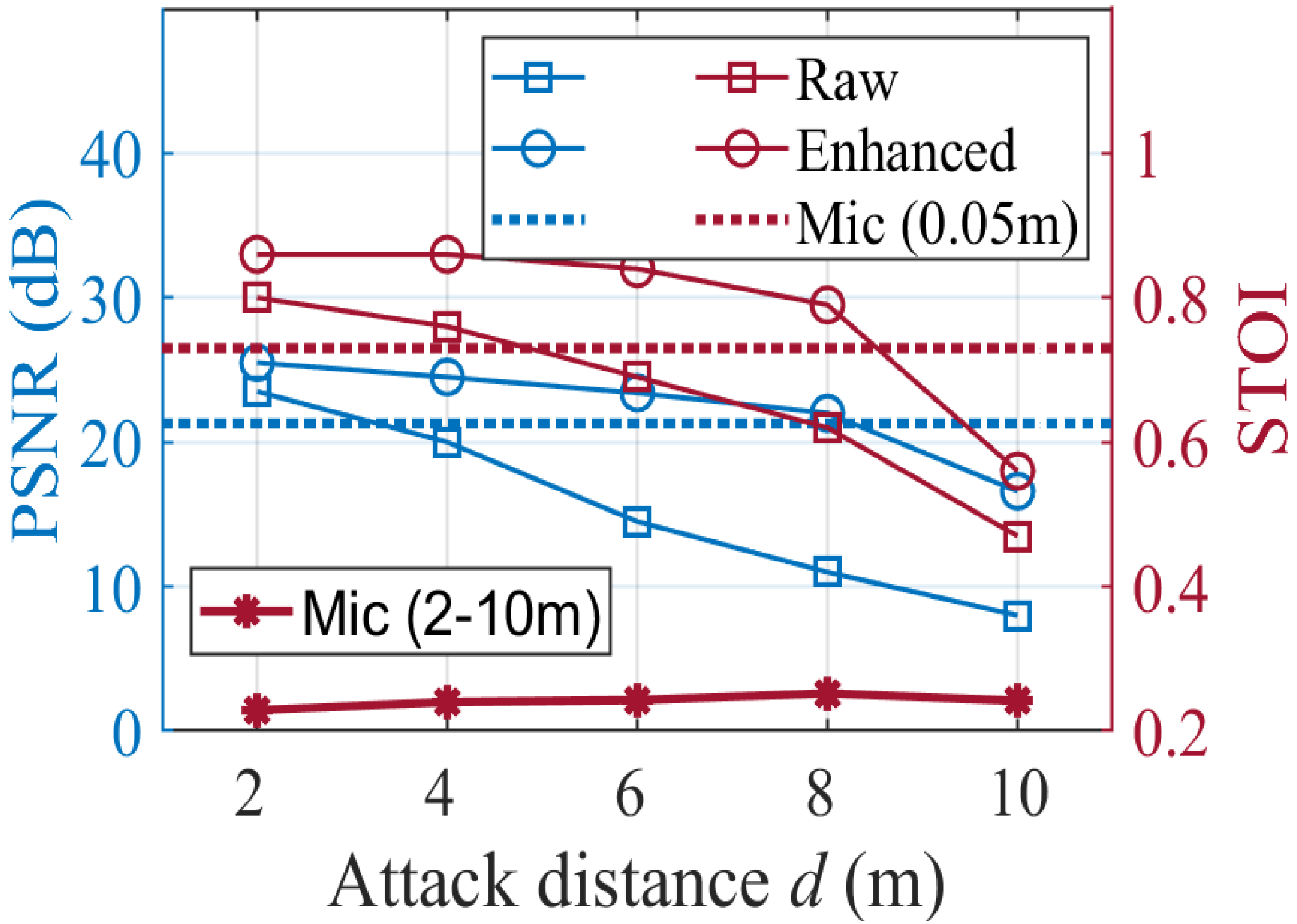
<u>The **motion interferences** are suppressed and the **speech quality** is improved by mmEve.</u>



Speech recovery @ 6m

Attack distance

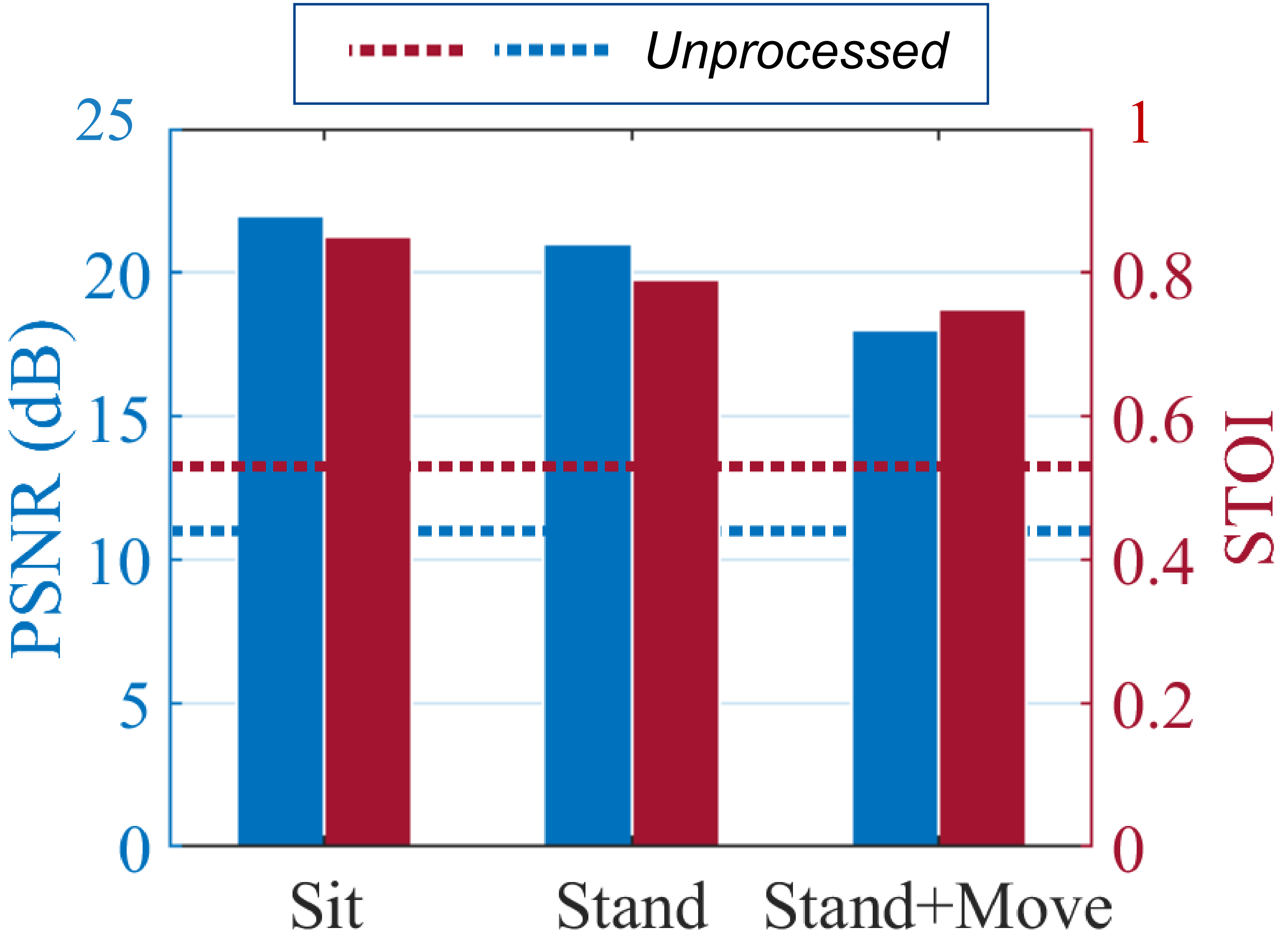
- Experimental setting
  - Distance: 2m~10m
  - Laboratory
- Result
  - Distance ↗
    Performance ↘
- Performance @ 6m
  - PSNR > 30dB
  - STOI > 0.7



Handhold condition

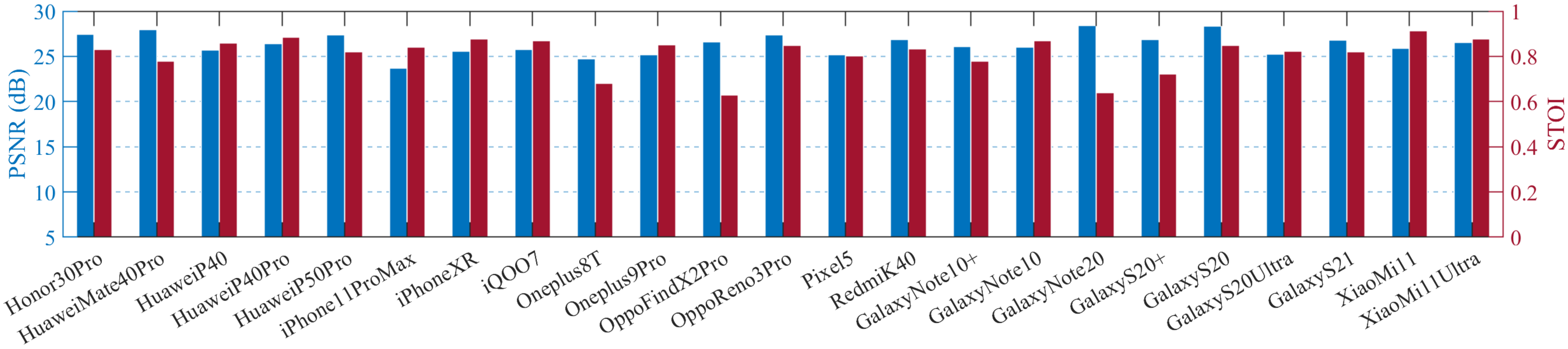
- Experimental setting
  - Sit on a chair
  - Stand and handhold
  - Stand and move

- Result
  - PSNR > 18dB
  - STOI > 0.75



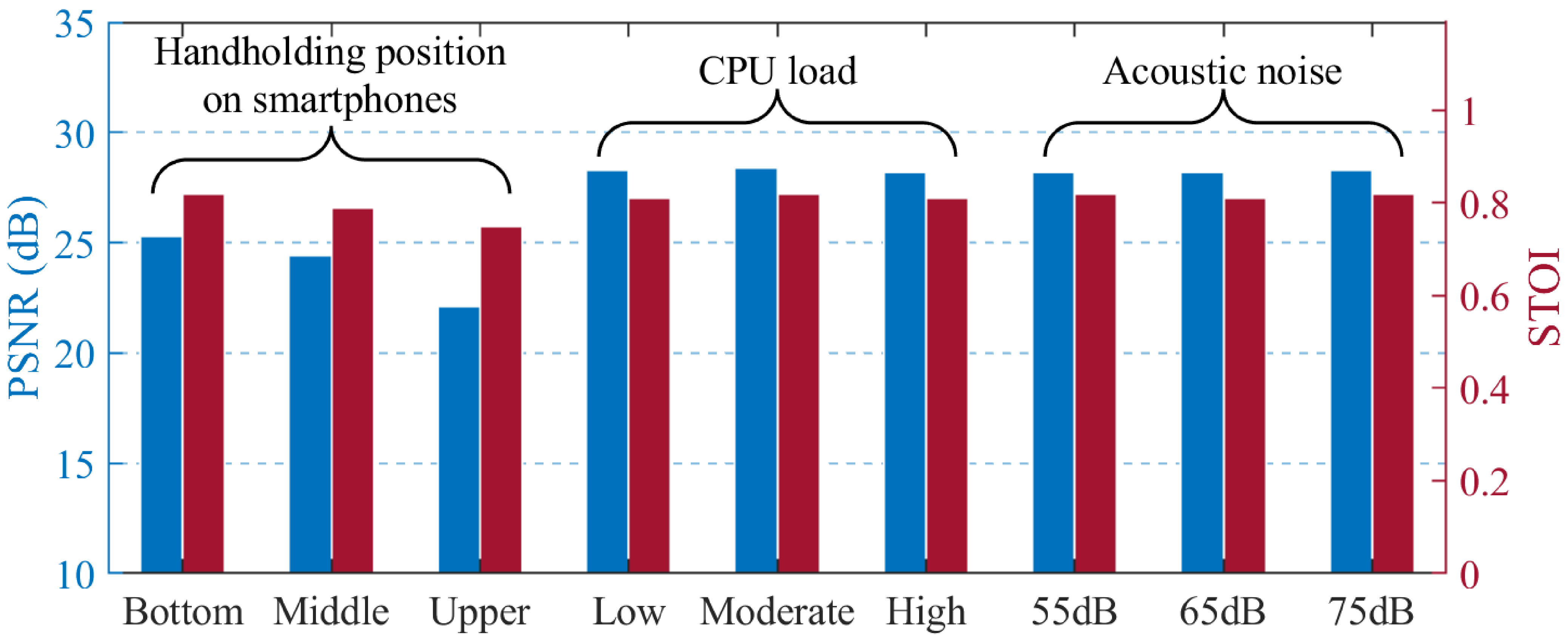
Different smartphones

- Twenty-three different smartphone models
  - Samsung, Huawei, Oppo, iPhone, etc.
- Result: PSNR > 18dB, STOI > 0.7



Complex condition

- Resilient to handholding habits / CPU load / acoustic noise



Outline

- Background
- Threat Model
- Feasibility Study
- System Design & Evaluation
- Defense
- Conclusion

Defense

- Active methods
  - Detect the malicious signals (77-81GHz) with sniffers
  - Jamming malicious devices
- Passive methods
  - Vibration damping (mitigate the vibration coupling)
  - Wave-absorbing materials (reduce the SNR of reflected signals)
  - Manipulate reflected signals with smart reflectors

Outline

- Background
- Threat Model
- Feasibility Study
- System Design & Evaluation
- Defense
- Conclusion

Conclusion

- We revealed a speech threat of smartphones posed by COTS mmWave sensors.
- We proposed an end-to-end system to recover audible speech from smartphone earpiece.
- We performed extensive experiments to investigate the threat level of the attack and gave the countermeasures.

Thanks for listening!