mmPhone: Acoustic Eavesdropping on Loudspeakers via mmWave-characterized Piezoelectric Effect

Chao Wang, Feng Lin, Tiantian Liu, Ziwei Liu, Yijie Shen, Zhongjie Ba, Li Lu, Wenyao Xu, Kui Ren

浙江大學
ZHEJIANG UNIVERSITY

University at Buffalo

Outline

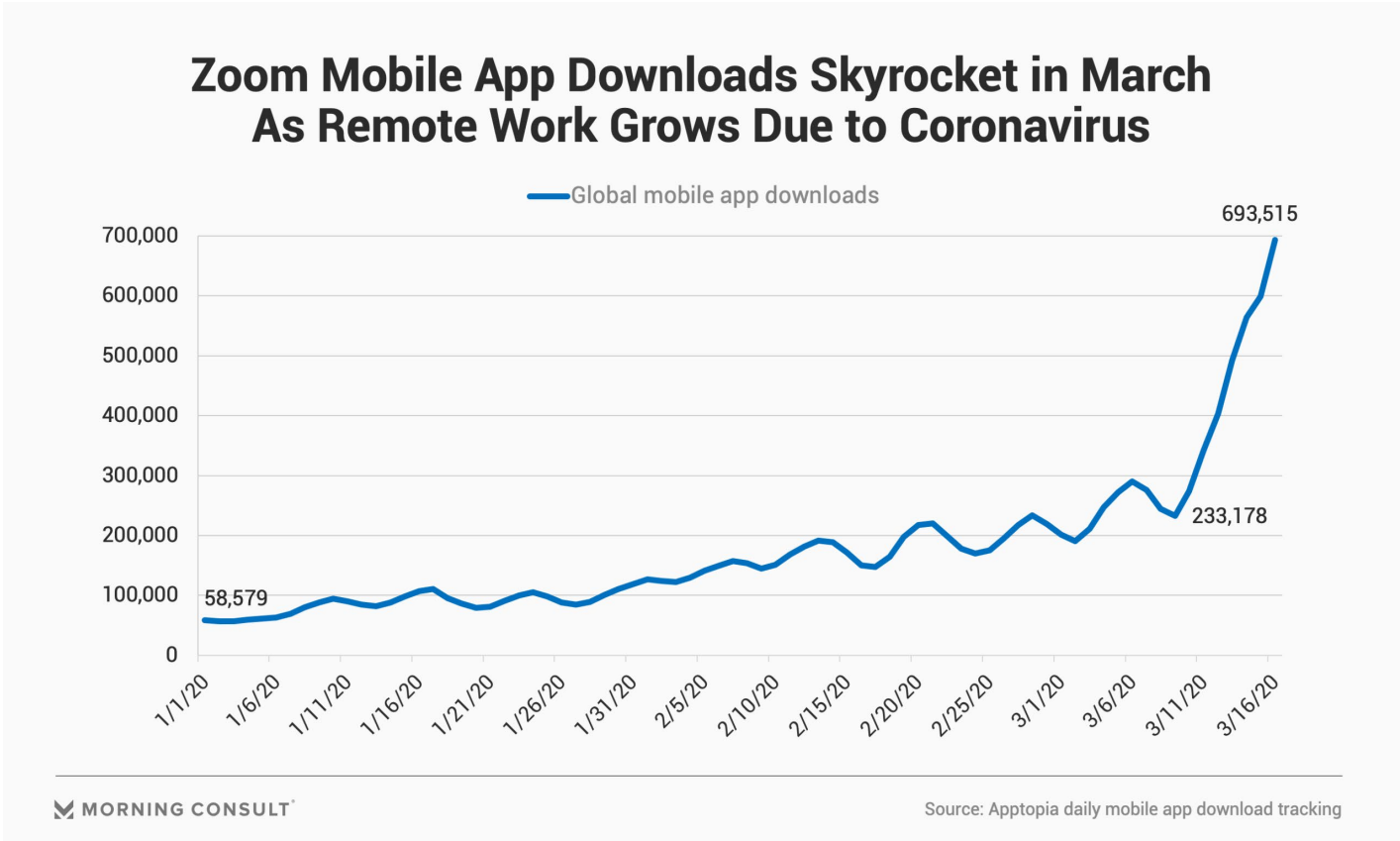
- Background
- Related Work
- Threat Model
- Sound-mmWave Transformation
- System Design & Evaluation
- Defense & Conclusion

Background

- Increasing demand of online voice communication

Video call

Virtual conference



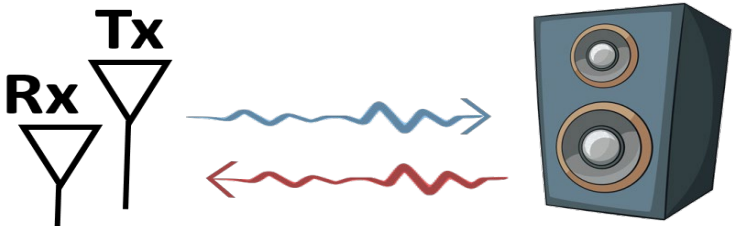
Sound Isolation

A user is participating an online conference in a soundproof room.

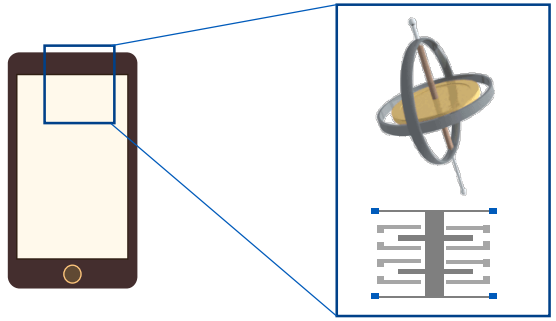
Soundproof rooms

Related work

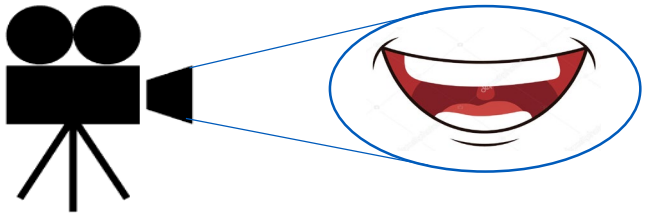
- Vibration-based eavesdropping
  - E.g., RF signals, motion sensors, video cameras, lidars…



RF signals (SenSys'20)



Motion sensors (NDSS'20)

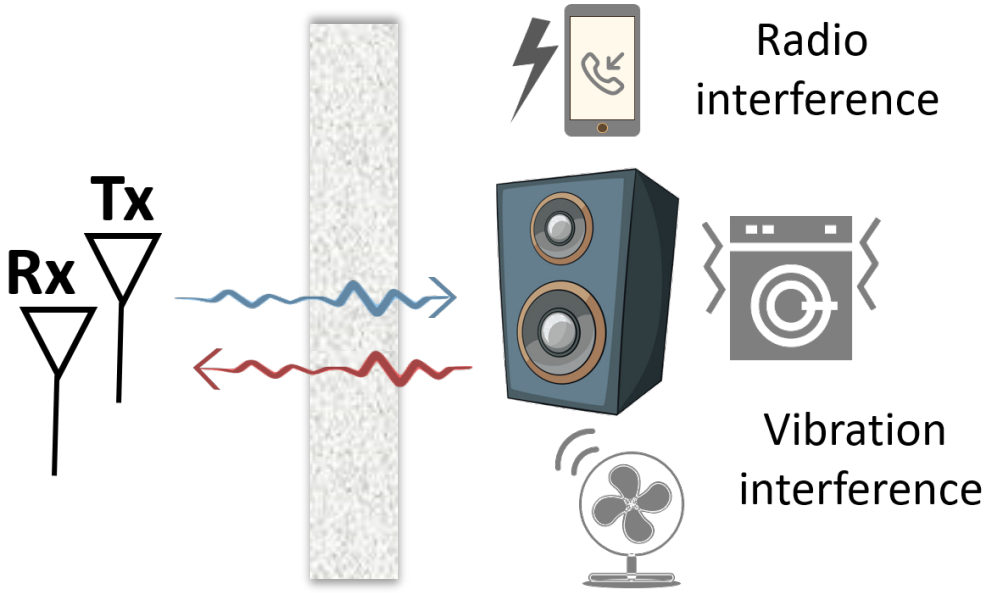


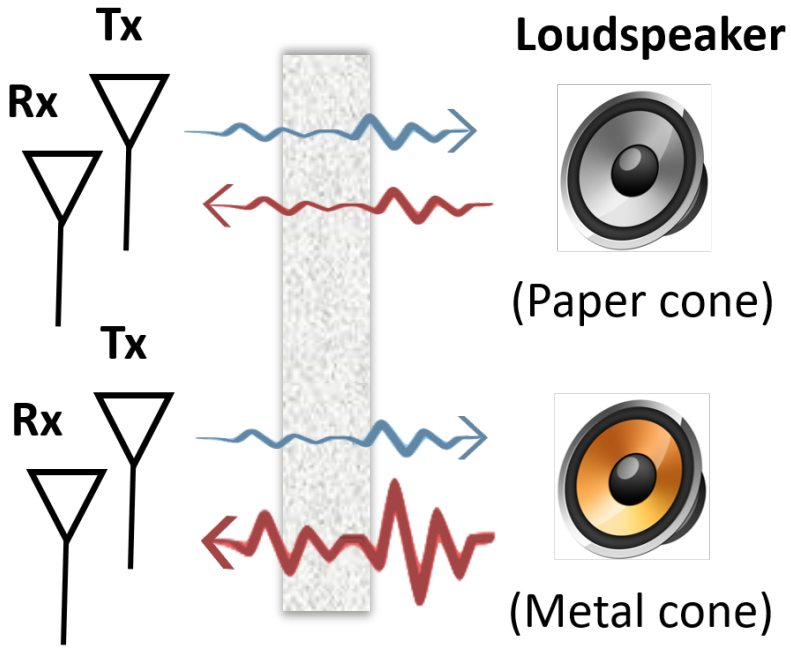
Video camera (SIGGRAPH'14)

Lidar sensors (SenSys'20)

Related work

- Wireless-based through-wall eavesdropping
  - Unrelated vibrating objects
  - Materials of targeted vibrating objects
  - Sound pressure level

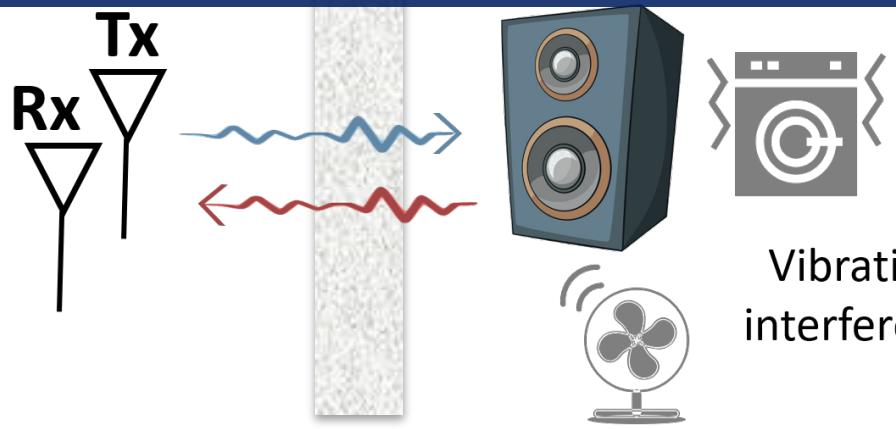




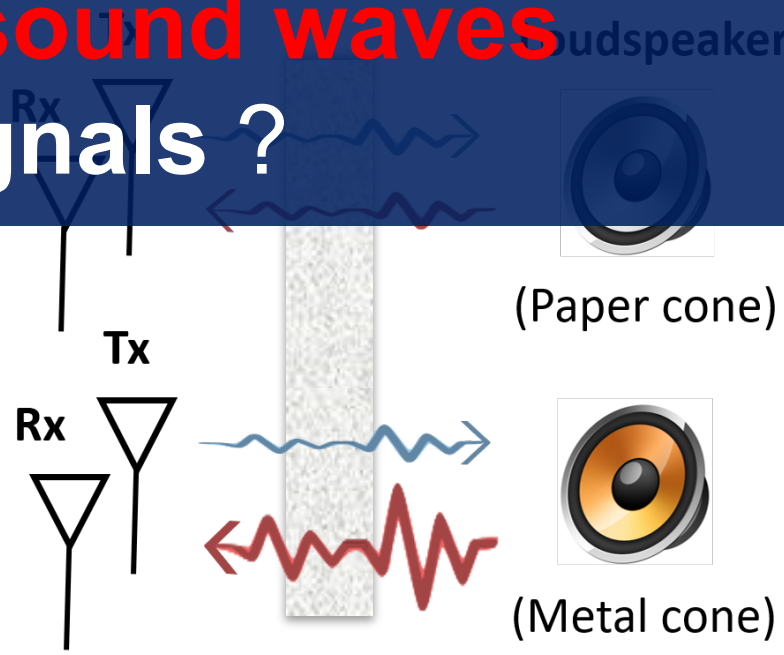
Related work

- Wireless-based through-wall eavesdropping
  - Unrelated vibrating objects
  - Materials of targeted vibrating objects
  - Sound pressure level

Recover propagating **sound waves** **via wireless signals** ?



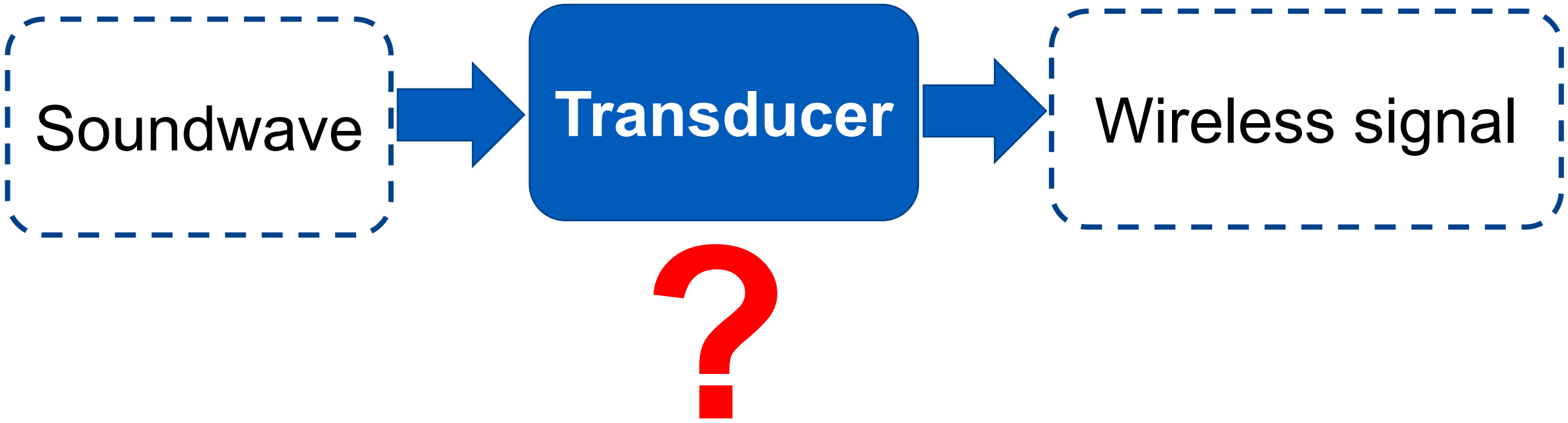
Vibration interference



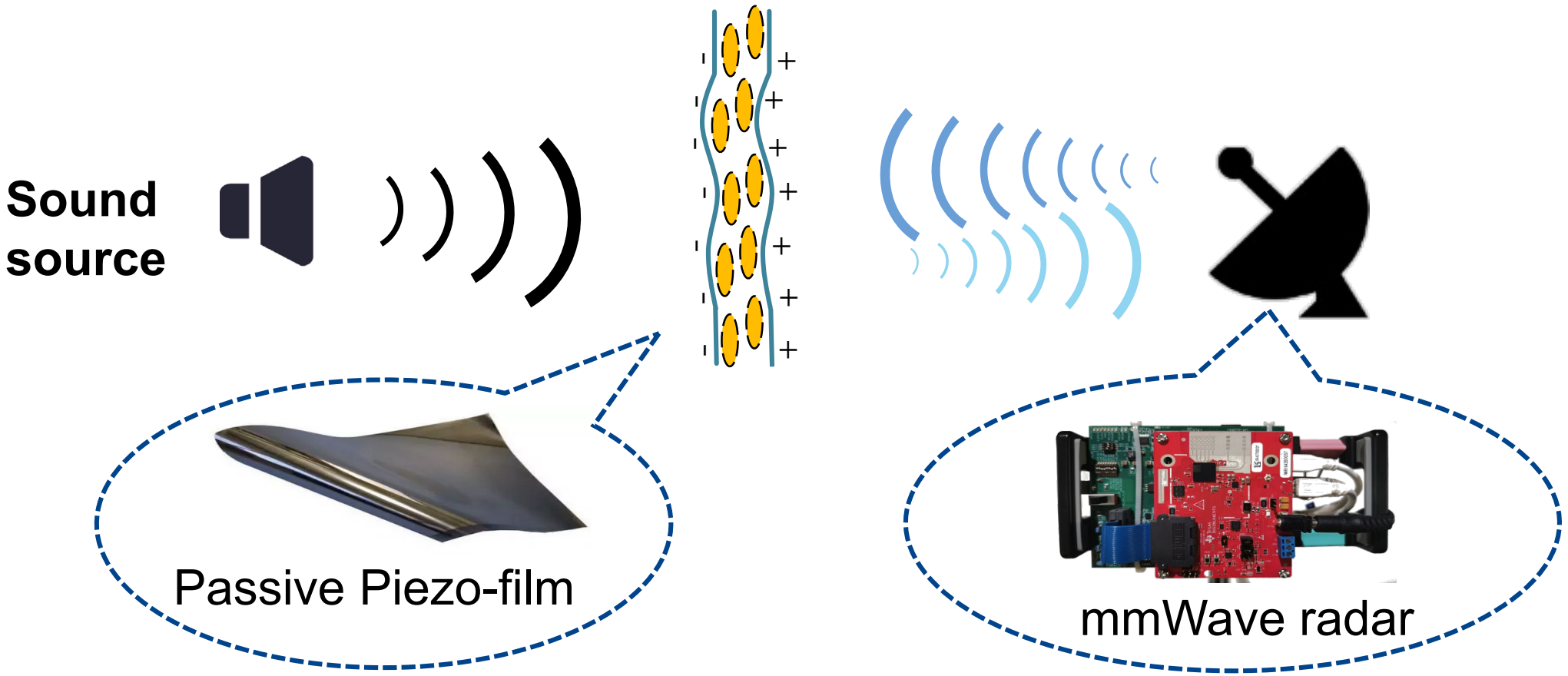
(Paper cone)

(Metal cone)

Sound-mmWave Transformation

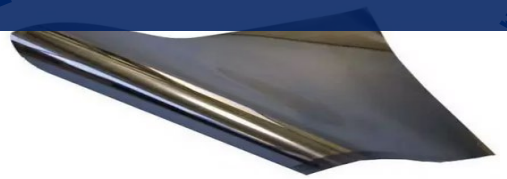


Sound-mmWave Transformation

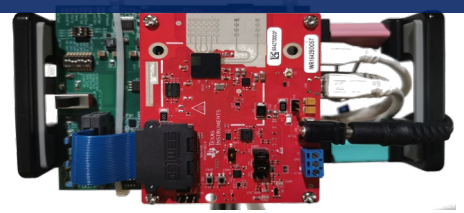


Sound-mmWave Transformation

Soundwaves can induce changes on the phase of **mmWave** signals reflected from the **piezo-film**



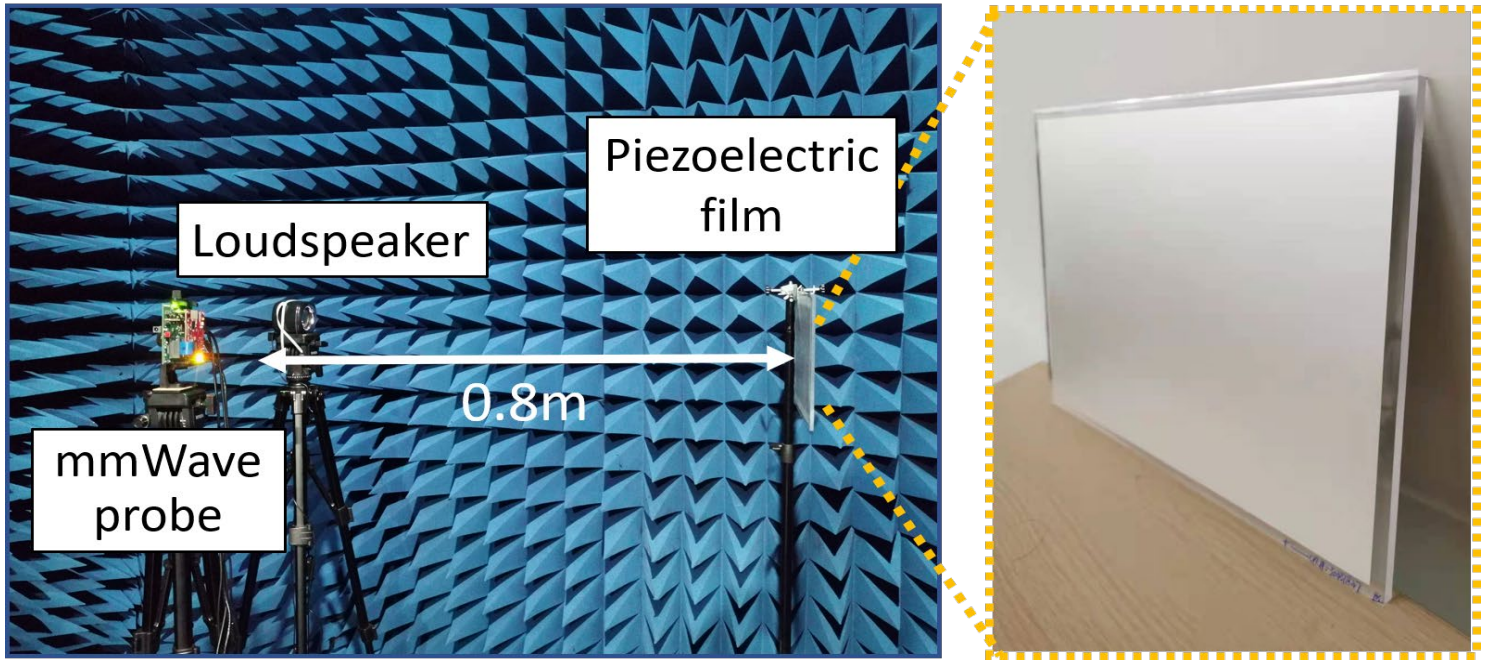
Passive Piezo-film



mmWave radar

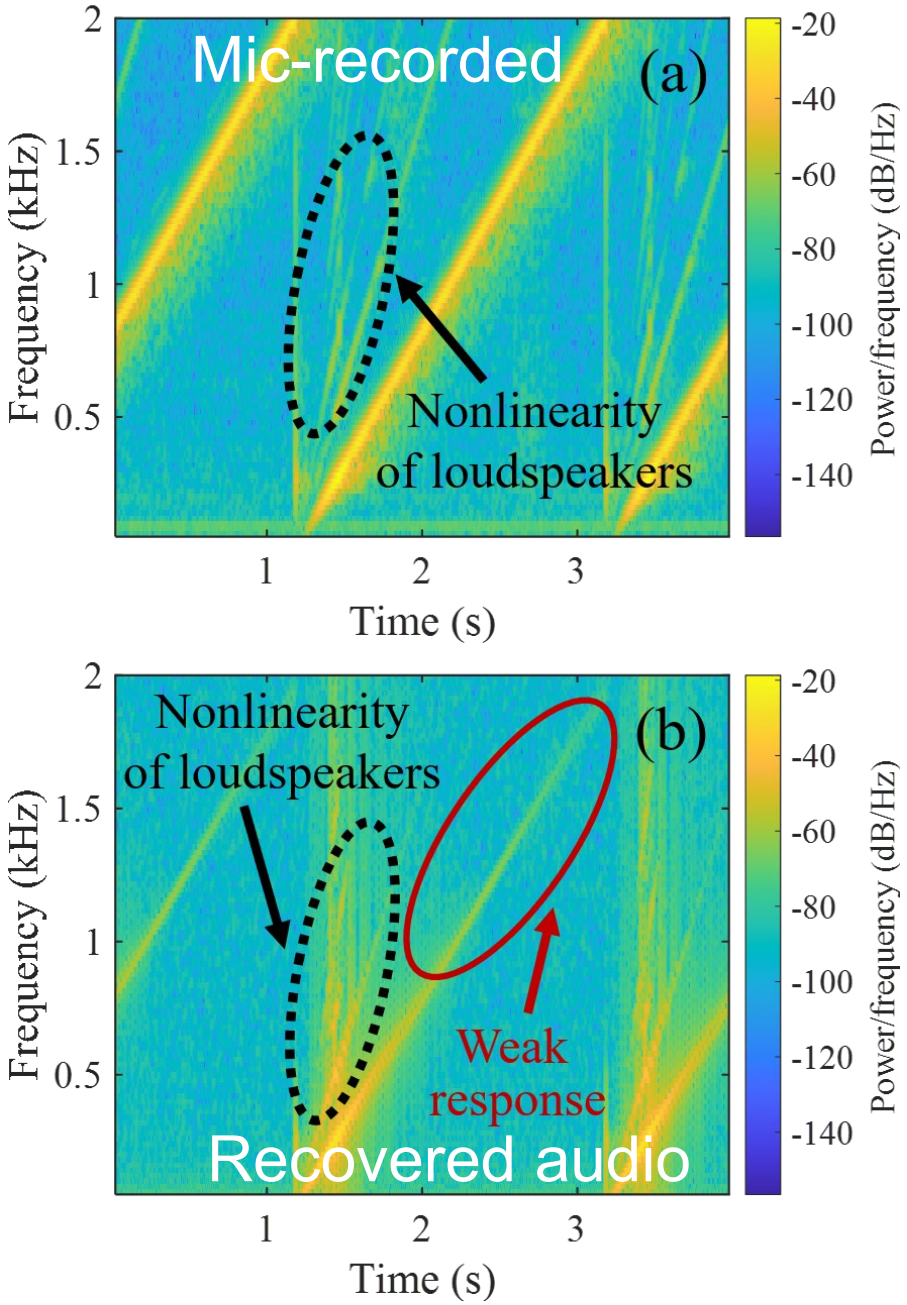
Feasibility Study

- Experiments in an anechoic chamber (LoS)
- **No physical connection** between the speaker and the probe
- **No physical vibration** on the film (stuck to a acrylic board)



Result

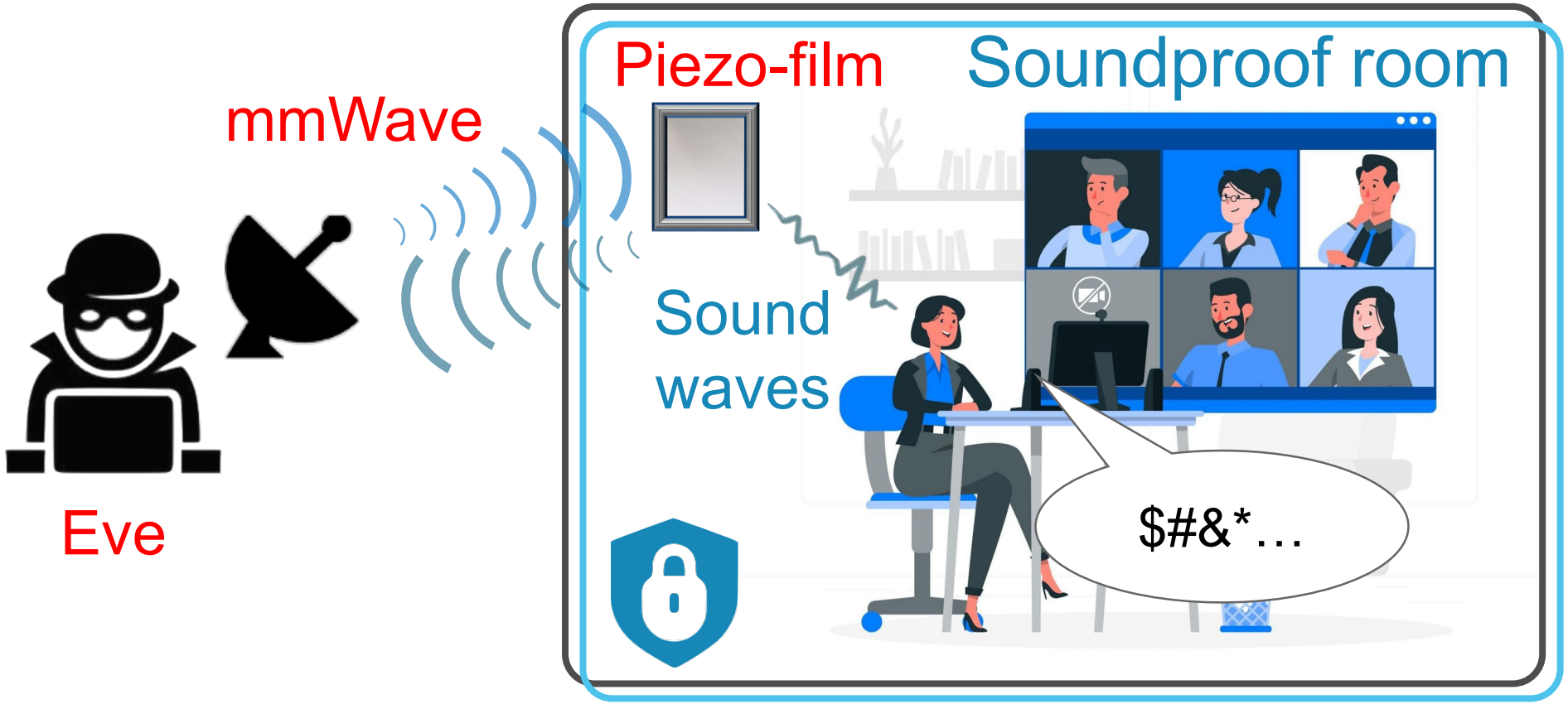
- Mic-recorded audio
  - Audible chirp
  - 0-2kHz
- mmWave-recovered audio
  - Audible chirp
  - 0-2kHz
  - Weak response in 1k-2kHz



Threat model

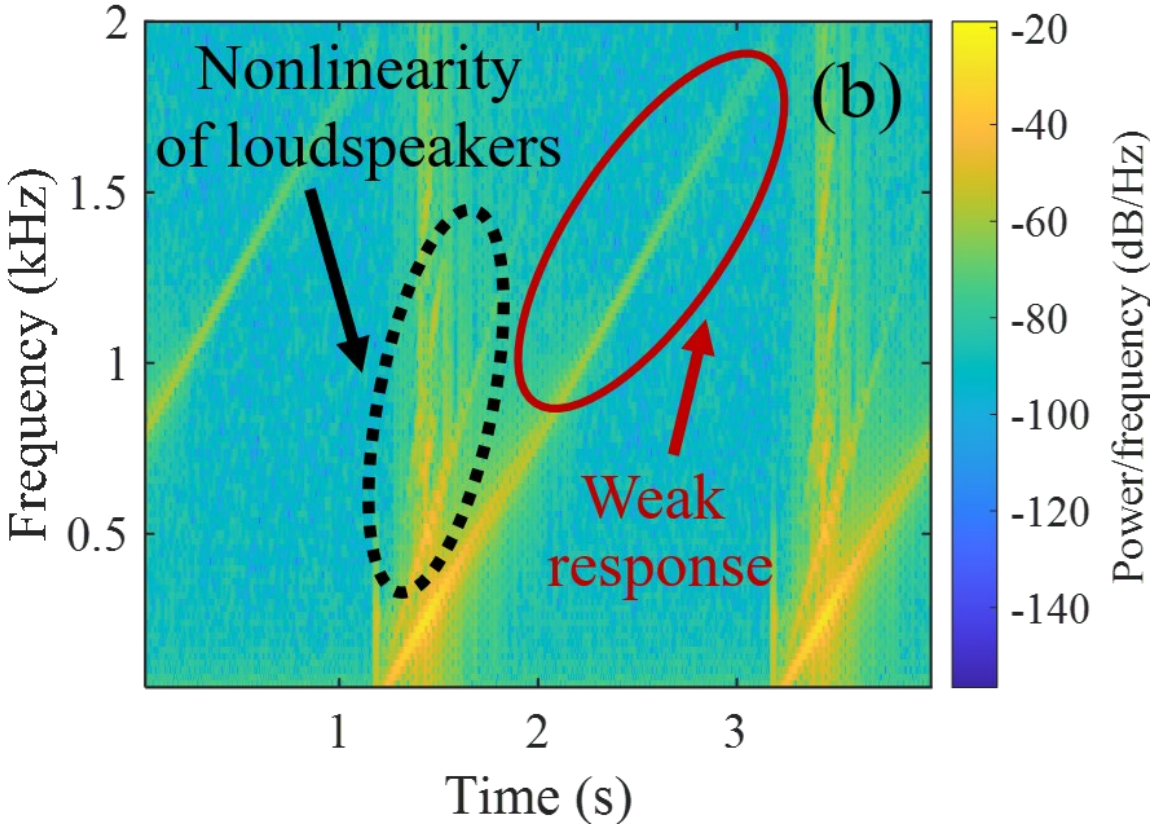
- Attack scenario
  - Soundproof
  - No active components
    - Mic or electronic devices

A new side channel via cross-modal perception



Findings in the feasibility study

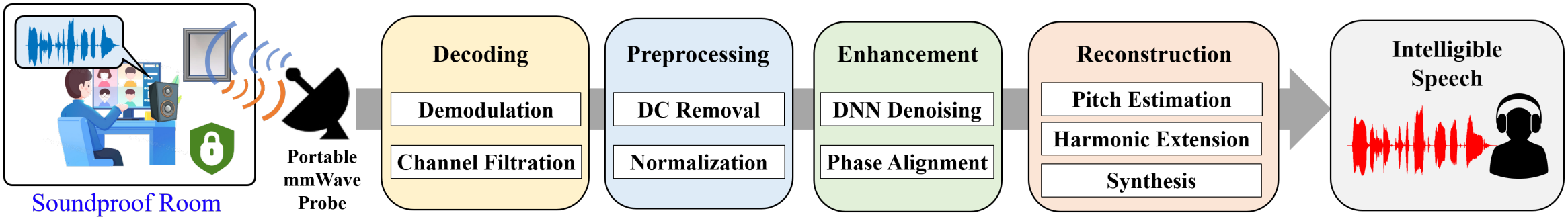
- Low power density
  - Long-range attack?
  - Through-wall attack?
- Weak response in 1k~2kHz
  - Loss of speech formants
  - Poor speech intelligibility



Recovered audio by the mmWave device

System Design

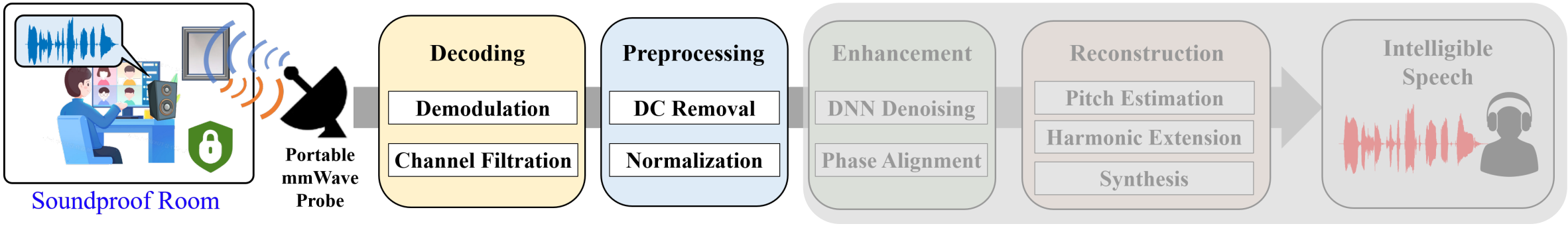
- **mmPhone: an end-to-end attack system**
  - Remote and **through-wall** eavesdropping
  - **High quality** and **intelligibility** speech recovery



mmPhone overview

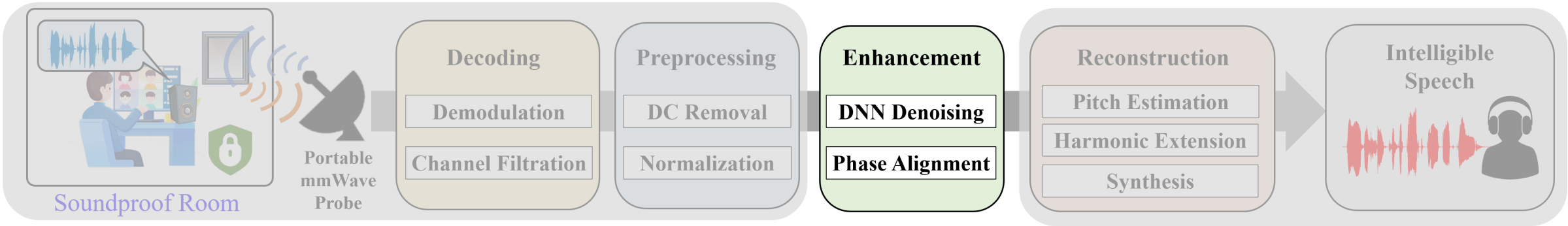
Decoding and Preprocessing

- Filtering out channels that contains speech
  - Band-pass filter (f_{c1}=80Hz, f_{c2}=250Hz)
  - Top-3 channels with highest power density are selected.
- Normalization
  - Constrain audio amplitude within [-1,1]



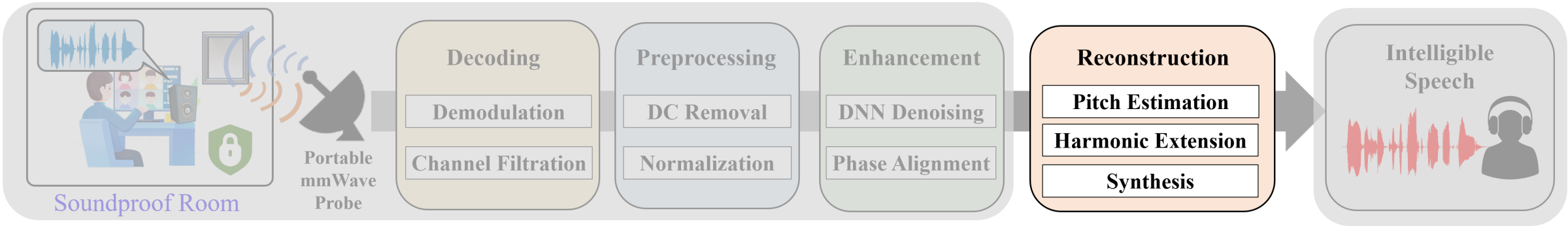
Enhancement

- Denoising Neural Network
  - Spectral mask estimation
- Enhance speech with multiple channels
  - Choose a baseline channel Ch_0
  - Align the phase of other channels with Ch_0



Reconstruction

- Each Rx chain can output an enhanced speech sequence
- Calibrated pitch estimation （85~255Hz）
  - Estimate pitch f_0^i for Antenna i $(i = 1,2,3,4)$
  - Calibrated pitch: $f_0 = \frac{\sum_{i=1}^{4} SNR_i \times f_0^i}{\sum_{i=1}^{4} SNR_i}$

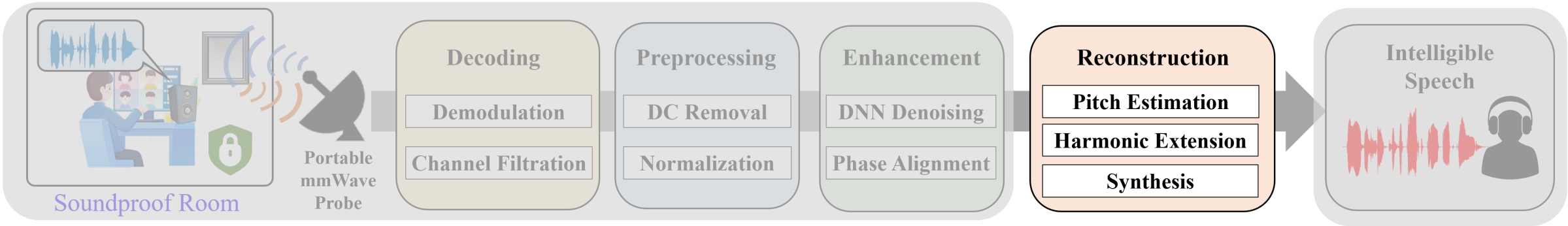


Reconstruction

- Spectral envelope estimation [1] + harmonic extension
- Synthesis (D4C algorithm [2])

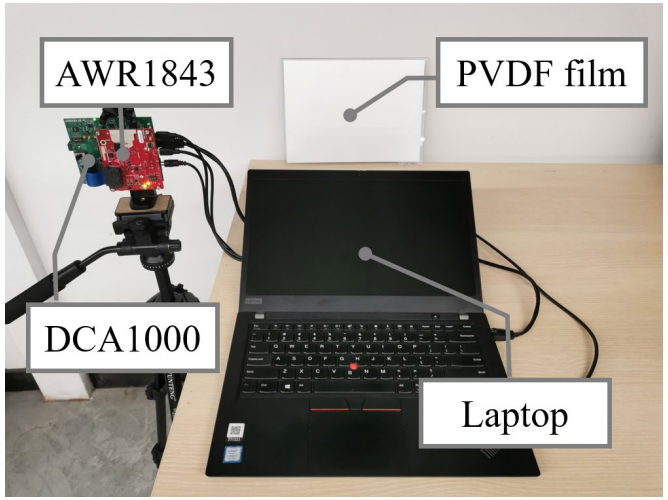
[1] M. Morise, “Cheaptrick, a spectral envelope estimator for high-quality speech synthesis,” Speech Communication, vol. 67, pp. 1–7, 2015

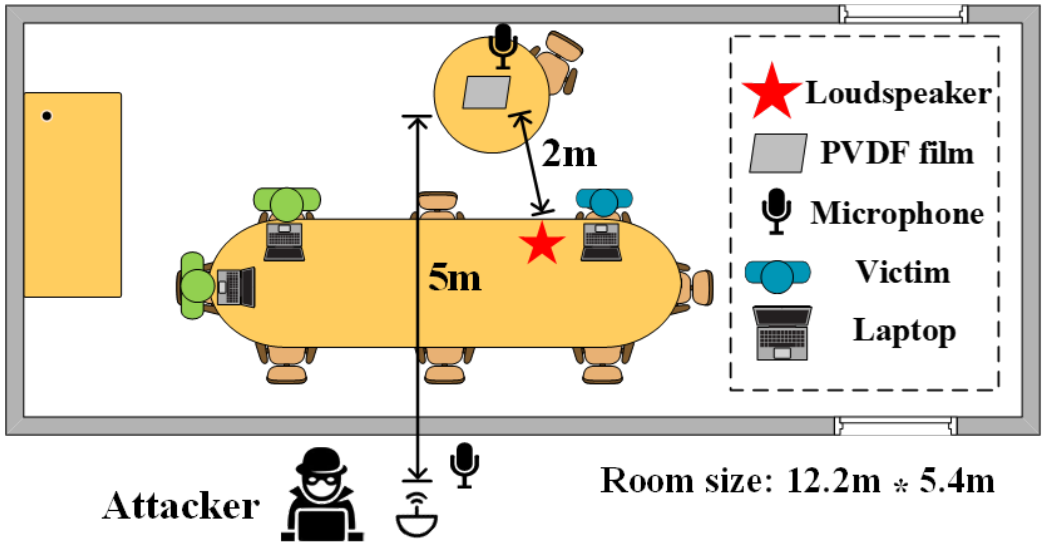
[2] M. Morise, “D4C, a band-aperiodicity estimator for high-quality speech synthesis,” Speech Communication, vol. 84, pp. 57–65, 2016.



Evaluation

- System setup
  - mmWave probe (AWR1843Boost)+ Piezo-film
  - Laptop (Thinkpad 490) + Server (GeForce RTX 2060 GPU)
- Conference room with soundproof glasses



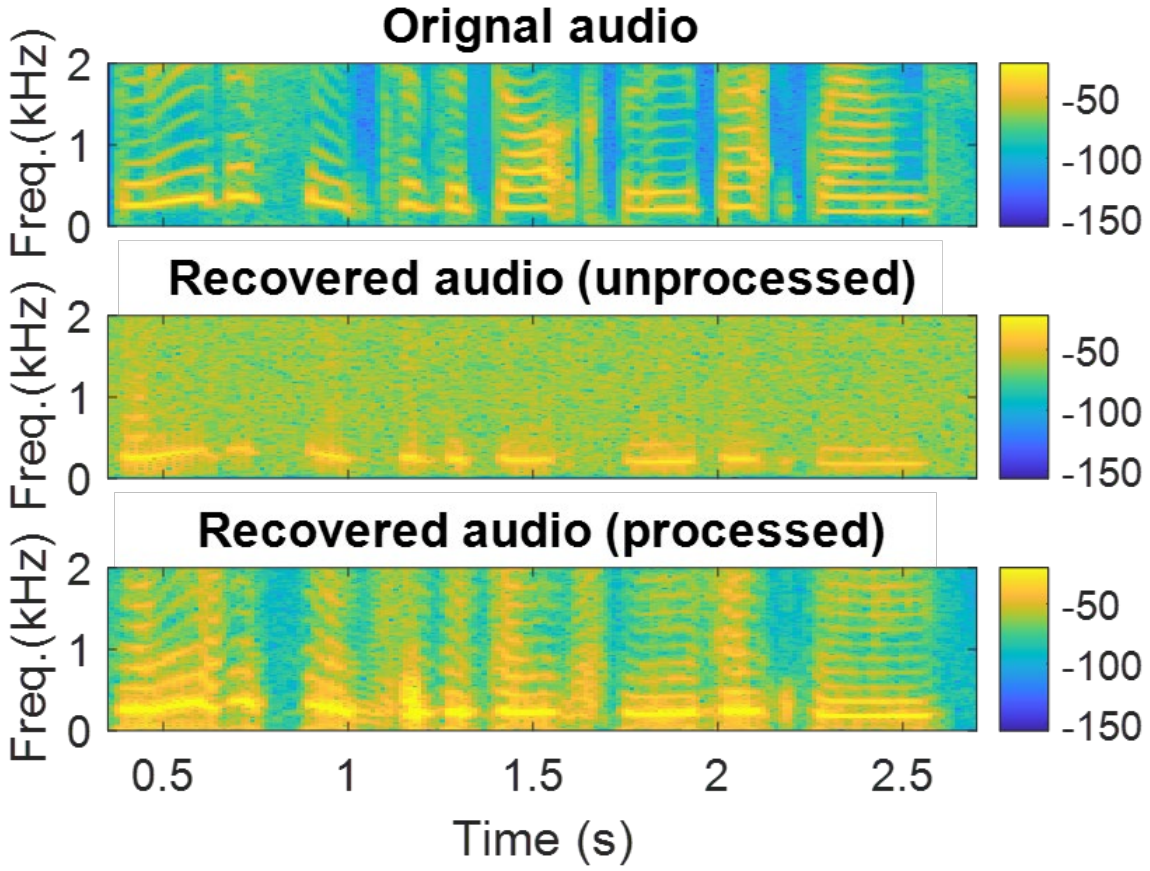


Evaluation

- Metric
  - Peak Signal-to-Noise Ratio (PSNR) : quantify the speech quality
  - Short-time Objective Intelligibility (STOI): quantify the speech intelligibility
- Dataset
  - Harvard Speech Corpus (HSC): 720 sentences
  - AudioMNIST: 10 digits from 60 speakers
  - Open Speech Repository (OSR): 100 sentences

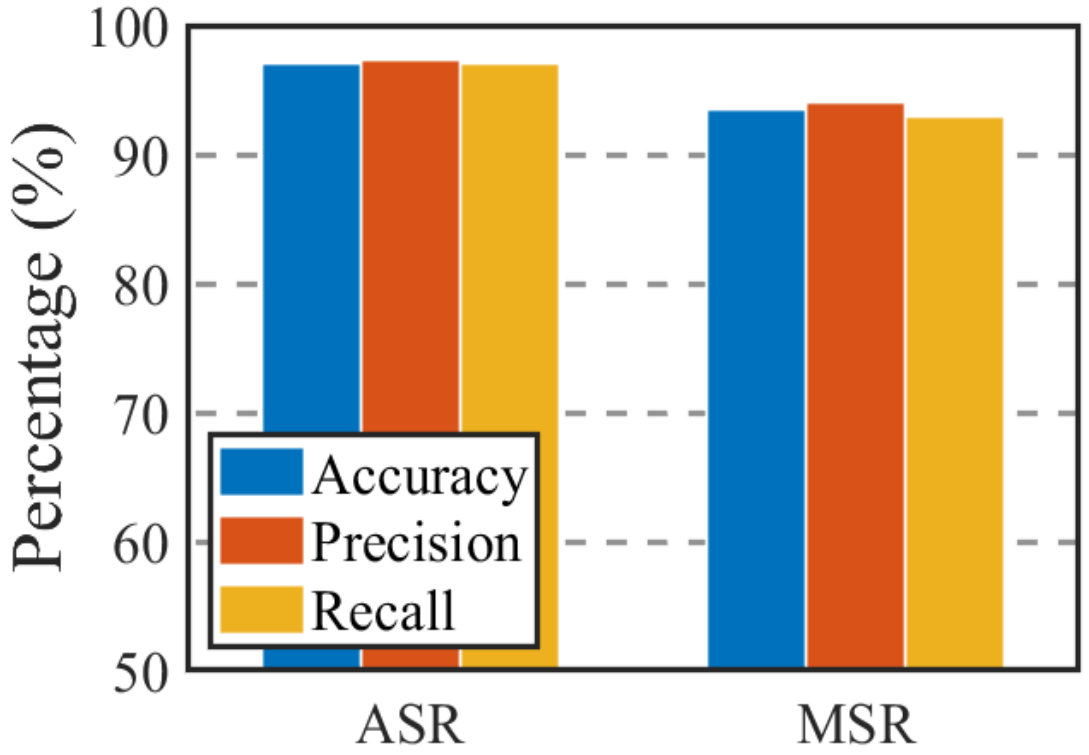
Sound recovery

- **Intelligible speech**
  - With a bandwidth up to 2.2kHz
- **High quality**
  - With little noise interference
- **Remote + through-wall**
  - Over 5m
  - Penetrating soundproof blockages



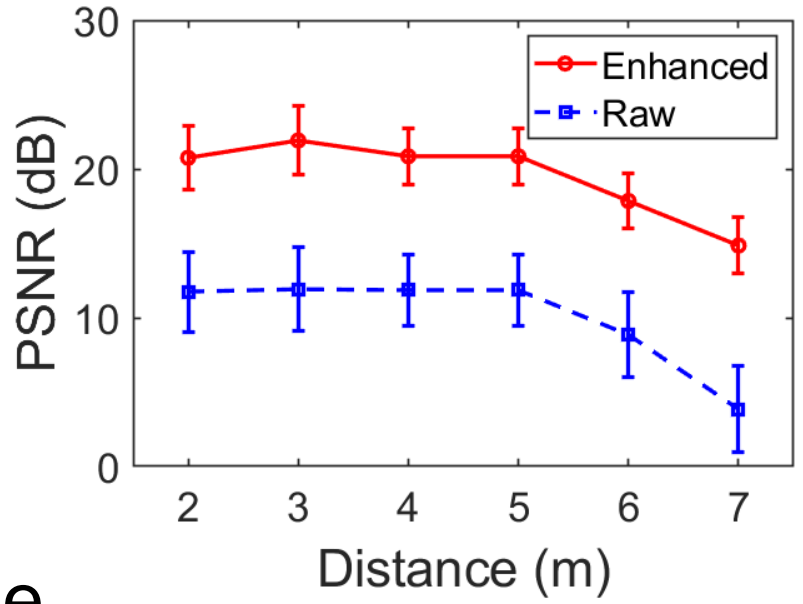
Digit recognition

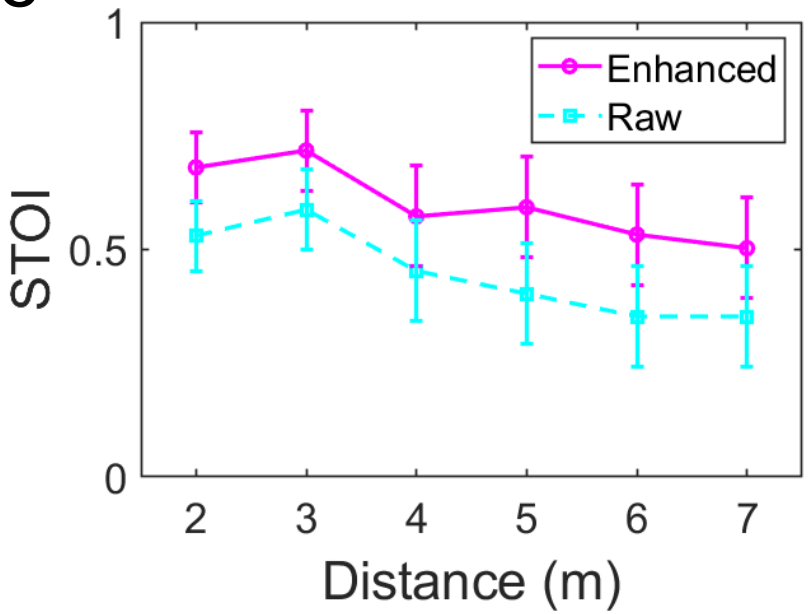
- Automatic Speech Recognition
  - Recognition model: ResNet-50
- Manual Speech Recognition
  - 15 volunteers
  - Recovered audio (0~9)
- Result
  - ASR: accuracy > 97%
  - MSR: accuracy > 93%



Attack distance

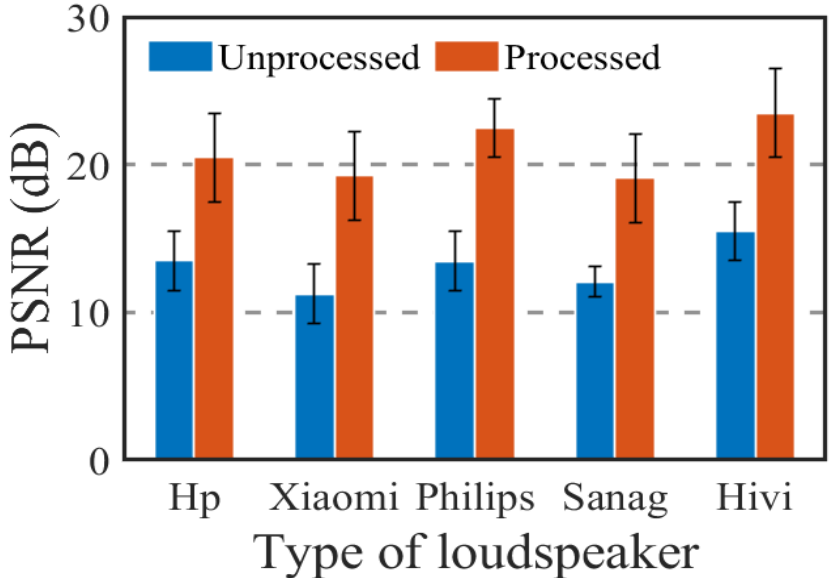
- Sensor-film distance
  - Over **5m**
- Raw recovered speech
  - Without processed by mmPhone
  - **PSNR>10dB, STOI>0.4**
- Enhanced speech
  - Processed by mmPhone
  - **PSNR>20dB, STOI>0.5**

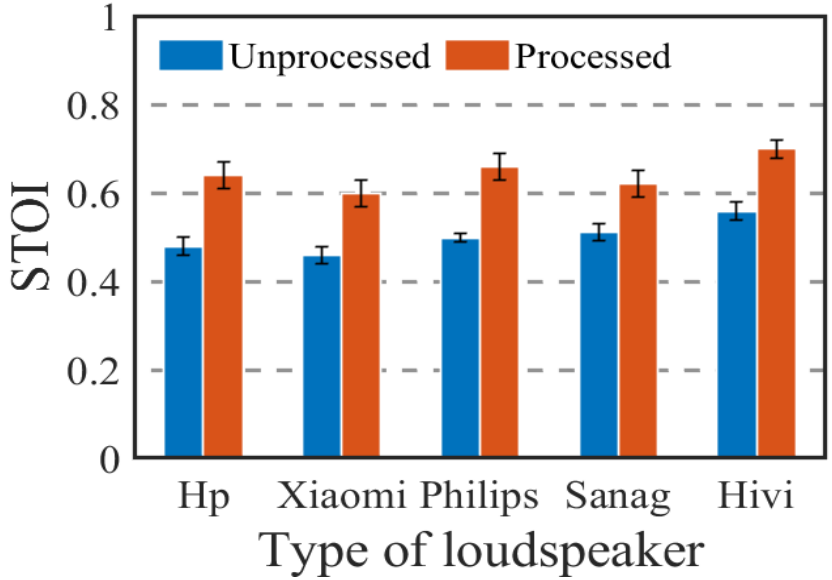




Different loudspeakers

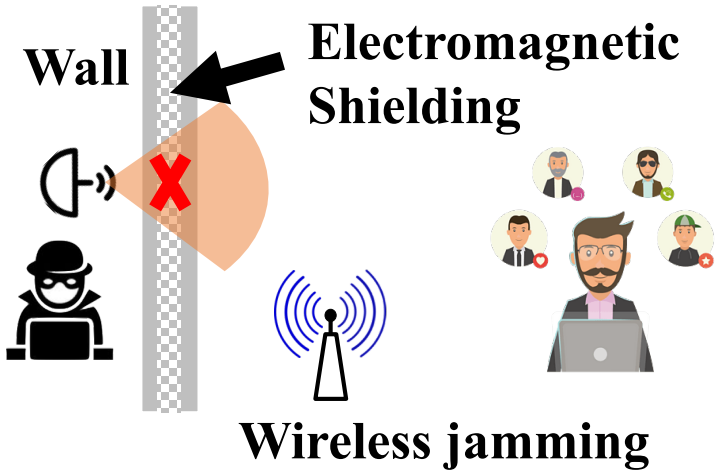
- Attack distance: 5m
- PSNR (19.1dB~23.5dB)
- STOI (0.59~0.70)





Countermeasures

- Blocking or interfering with the mmWave
  - Electromagnetic shielding
  - Jamming with mmWave signals
- Prevent the propagation of sound waves in the air
  - Wearing a headset or earphone



Conclusion

- A new cross-modal perception scheme
  - Recover sound waves (Mechanical Waves) with mmWaves (Electromagnetic Waves)
  - A new type of “microphone” via mmWave interrogation

- A new attack via mmWave-characterized piezo-effect
  - **Intelligible speech** with **high quality** can be recovered over 5m through the wall.
  - Soundproof protection is not reliable.

Thanks for listening!