RF-Parrot: Wireless Eavesdropping on Wired Audio

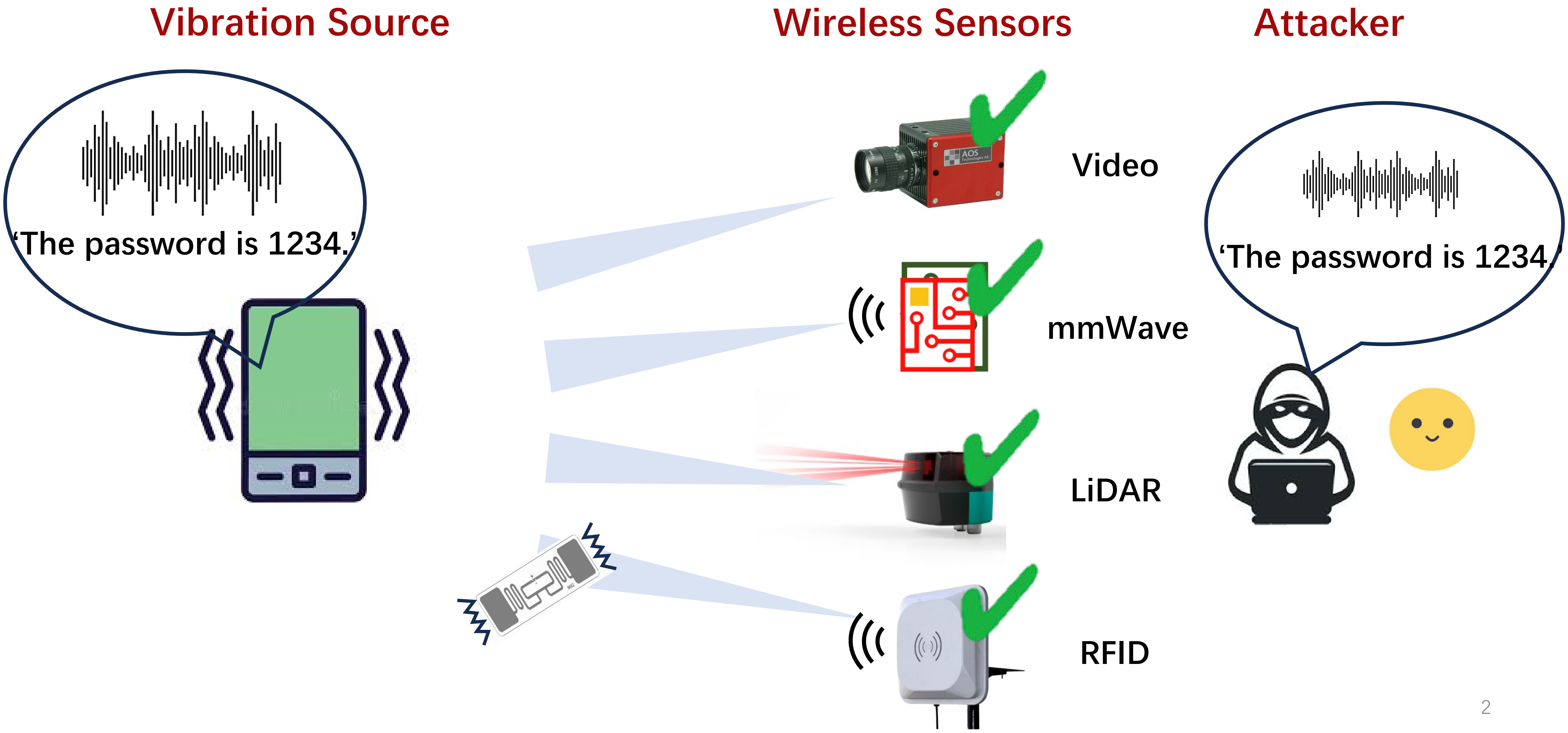
Yanni Yang*, **Genglin Wang†**, Zhenlin An‡, Guoming Zhang*, Xiuzhen Cheng*, Pengfei Hu*

*School of Computer Science and Technology, Shandong University, China

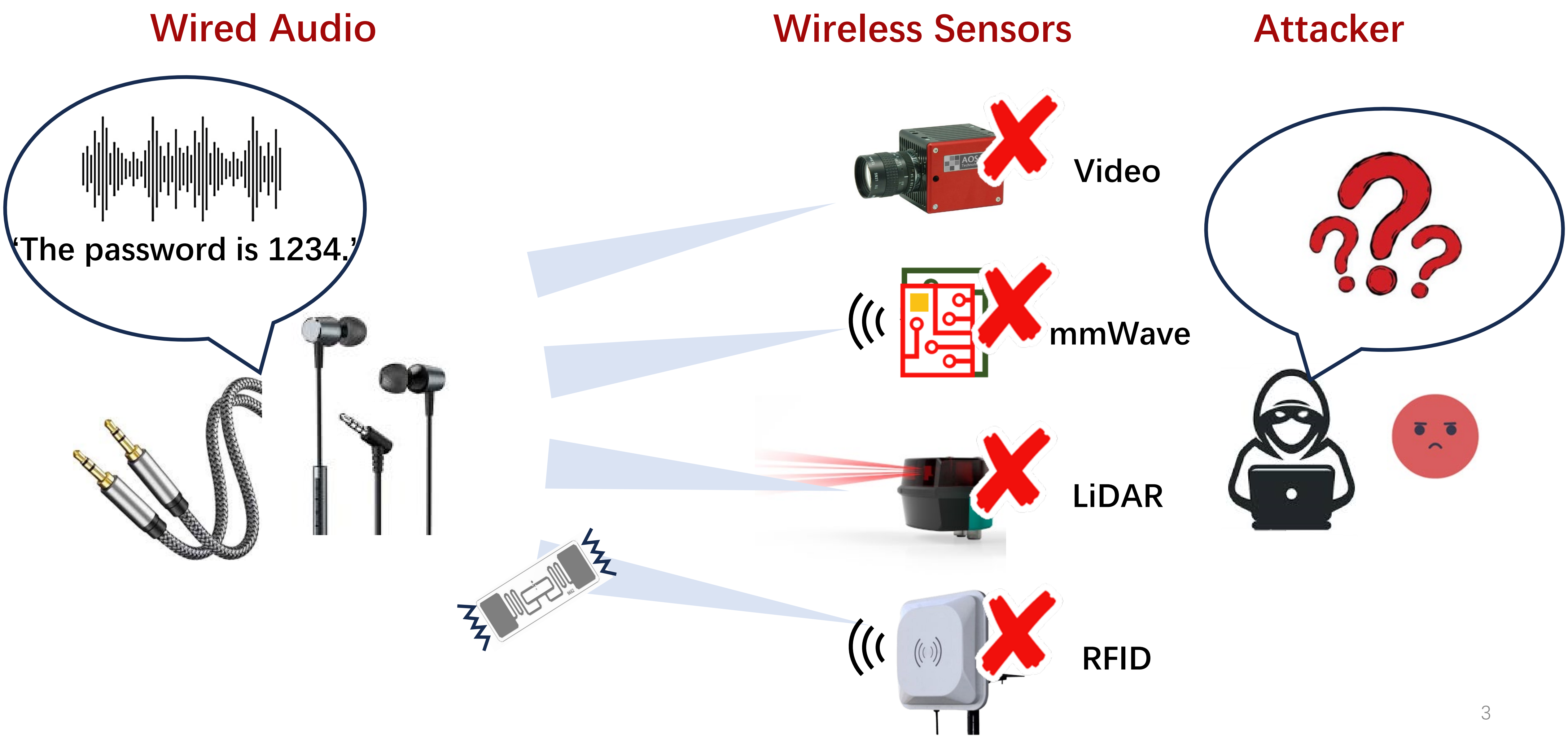
†School of Information Science and Engineering, Shandong University, China

‡Department of Computer Science, Princeton University, USA

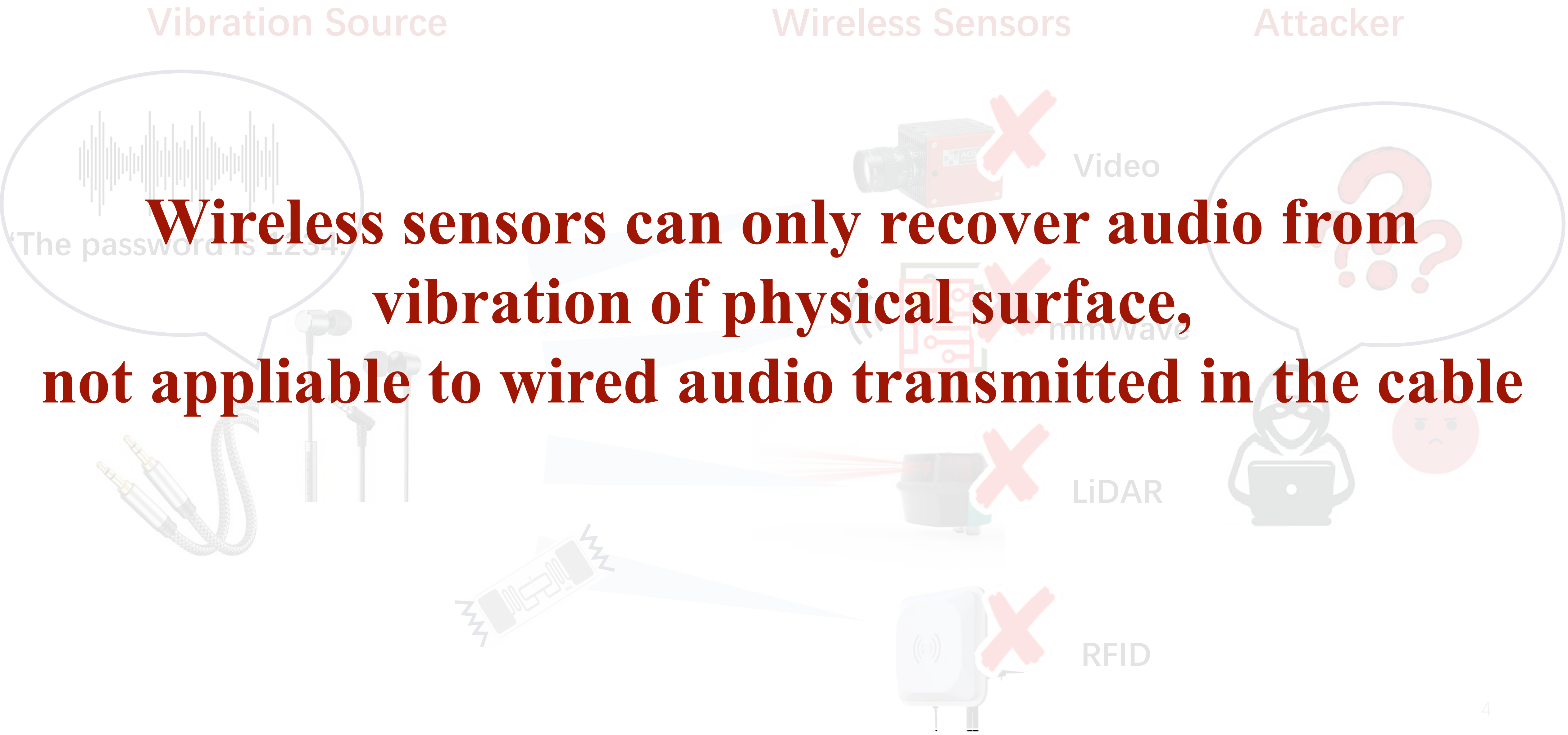
Introduction



Introduction



Introduction



Wireless sensors can only recover audio from vibration of physical surface,
not appliable to wired audio transmitted in the cable

Introduction

A possible solution: Radio-frequency Retroreflector Attack (RFRA)

RFRA was disclosed in ANT Catalog.

Michael Ossmann reproduces RFRA at DEF CON'22

Introduction

Related works about RFRA:

Existing RFRAs are only appliable to digital signals (e.g., PS/2, USB, VGA signals).

Surlyspawn (RF retroreflector) in ANT Catalog *Ragemaster (RF retroreflector) in ANT Catalog*

Question: Can we wirelessly eavesdrop on the wired analog audio with RFRA?

Michael Ossmann reproduces RFRA at DEF CON'22 *S Wakabayashi, et al. at WOOT'18*

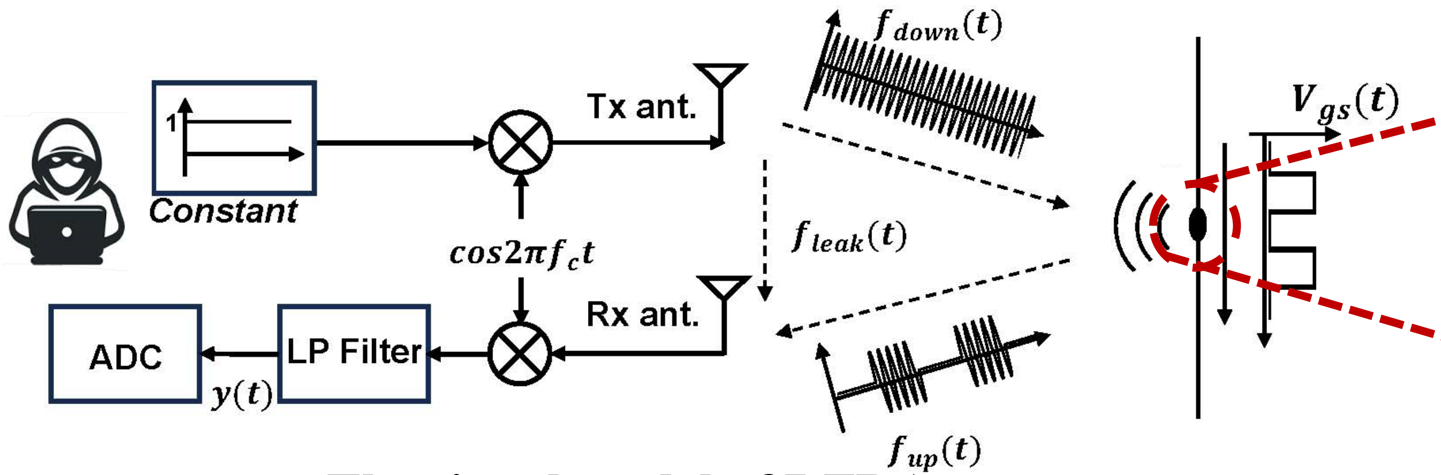
Introduction

Contributions:

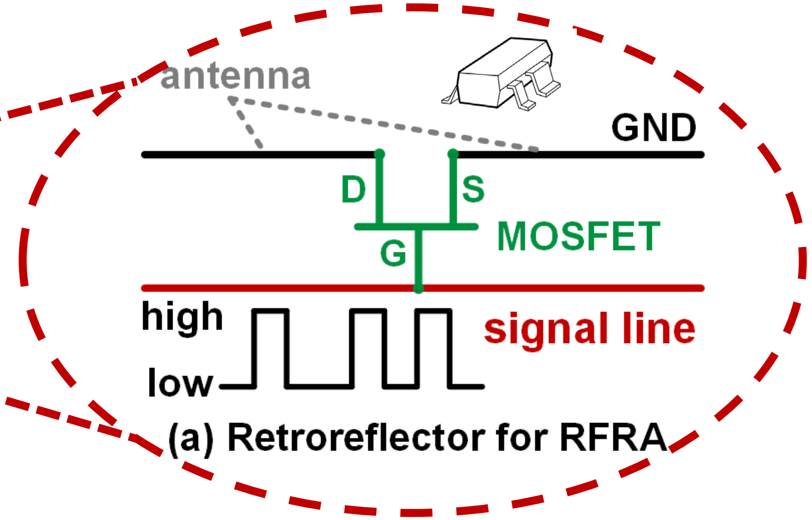
- We propose the first analog RFRA system, **RF-Parrot**, for eavesdropping on the wire-transmitted analog audio signal remotely via a new design of the retroreflector made from the D-MOSFET.
- We demonstrate that RF-Parrot can intercept analog audio signals at a distance of 1 m through the wall.
- We evaluate RF-Parrot using over 65,000 speech commands from thousands of people in various environments.

Preliminaries

Digital RFRA:



The signal model of RFRA



(a) Retroreflector for RFRA

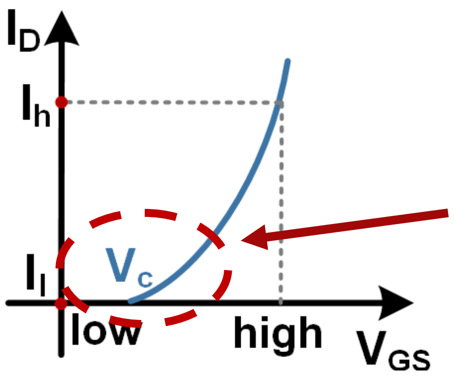
- ➢ Enhancement mode MOSFET (E-MOSFET) is the retroreflector.
- ➢ When the digital voltage signal $V_{gs}(t)$ varies, the E-MOSFET switches between reflective and non-reflective states, amplitude-modulating the reflective waves.

Preliminaries

Why digital RFRA does not work for analog audio?

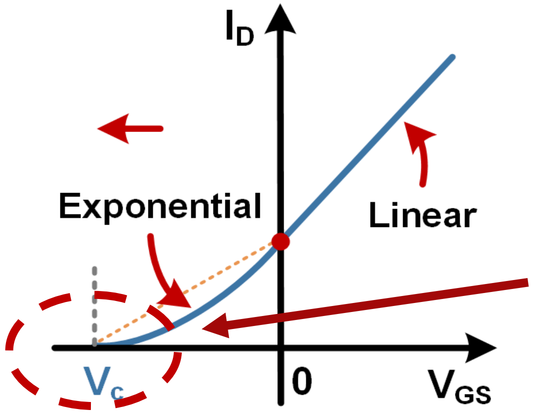
➢ **The E-MOSFET can only be activated by positive voltage signal.**

✗ The E-MOSFET's transfer curve:



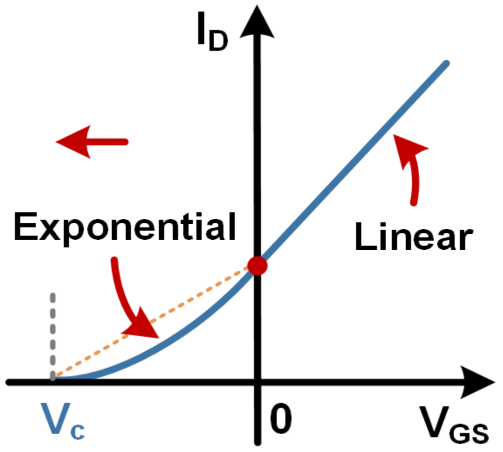
Positive cut-off voltage

✓ The Depletion mode MOSFET's (D-MOSFET) transfer curve:

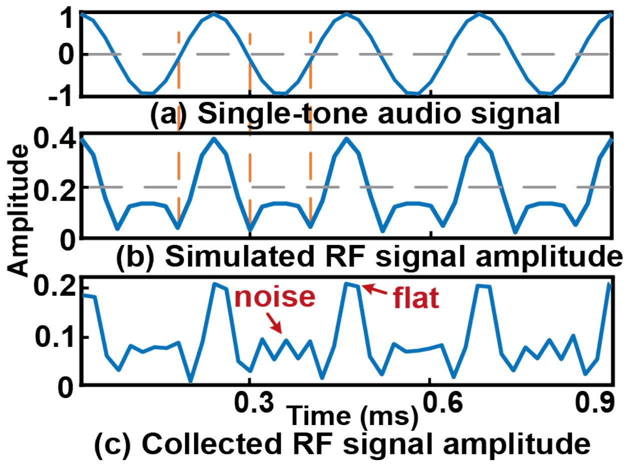


Negative cut-off voltage

Preliminaries



The D-MOSFET's transfer curve



Simulated & collected RF signal

The impact of D-MOSFET's transfer curve on received RF signal $y(t)$:

$$y(t) = \begin{cases} \alpha \cdot V_{gs} + \beta & V_{gs} \geq 0 \\ -e^{\gamma \cdot V_{gs}} \cdot V_{gs} & V_c \leq V_{gs} < 0 \\ 0 & V_{gs} < V_c \end{cases}$$

→ **(near) Linear.**

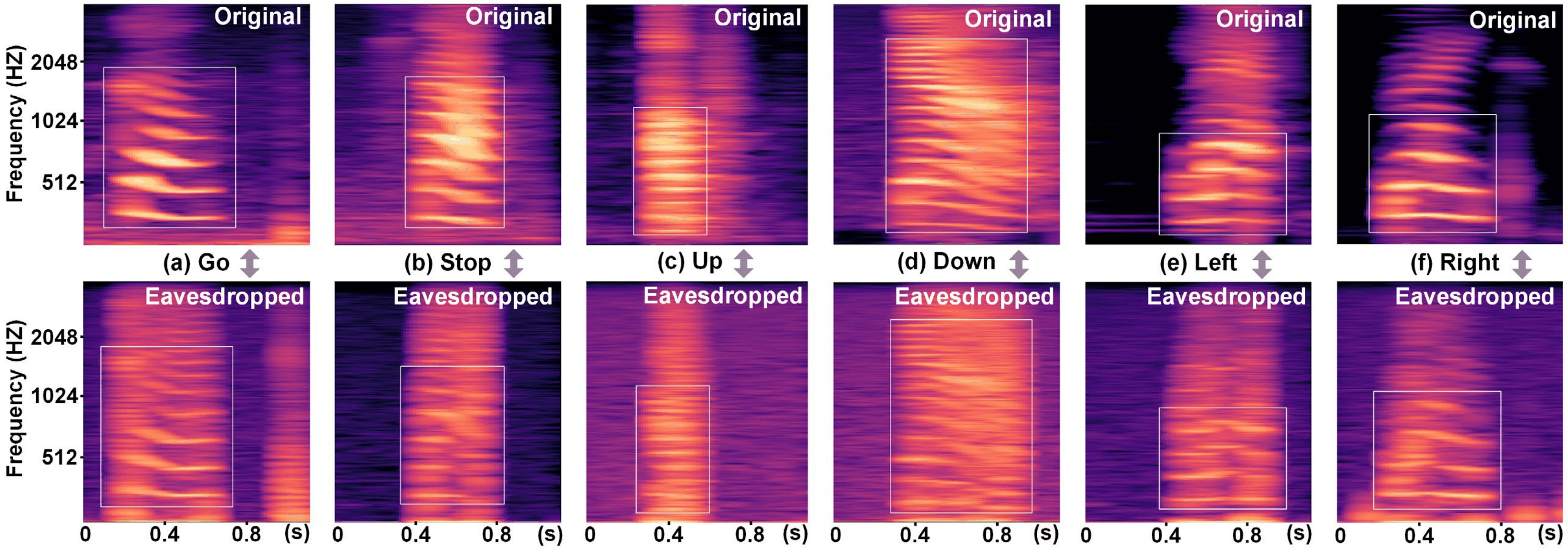
→ **The negative part kept, but distorted non-linearly.**

Preliminaries

The comparison between MCD values of E-MOSFET and D-MOSFET:

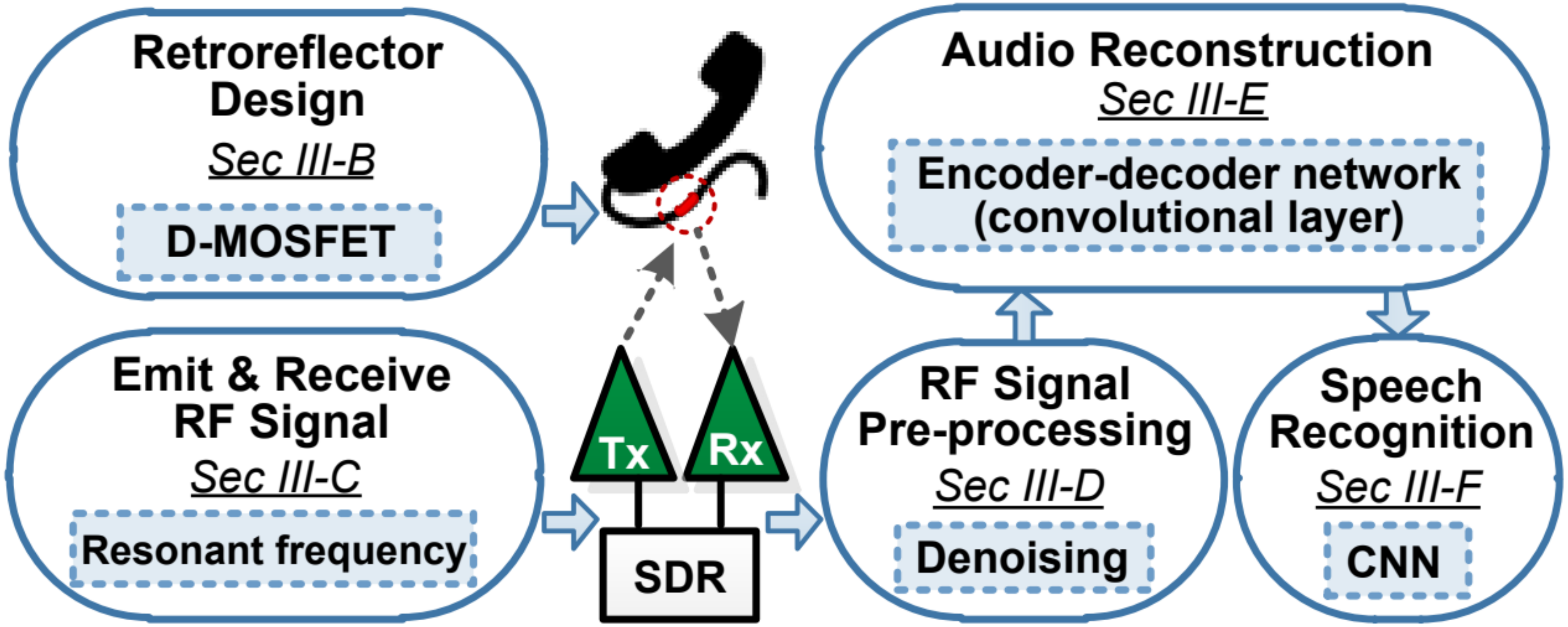
	0	1	2	3	4	5	6	7	8	9
D-MOS	16.9	17.0	16.8	17.7	17.0	17.4	17.7	18.4	16.2	18.5
E-MOS	20.6	23.4	23.9	22.9	22.3	22.6	22.1	24.2	25.9	25.4

The spectrograms of speech commands (go, stop, up, etc.)



System Design

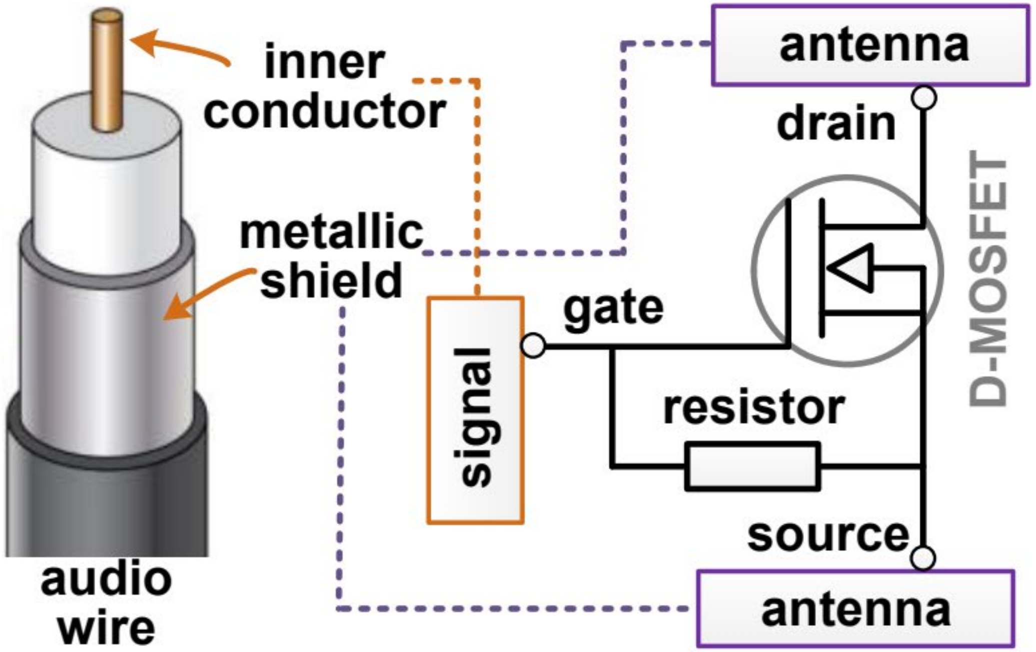
System Overview:



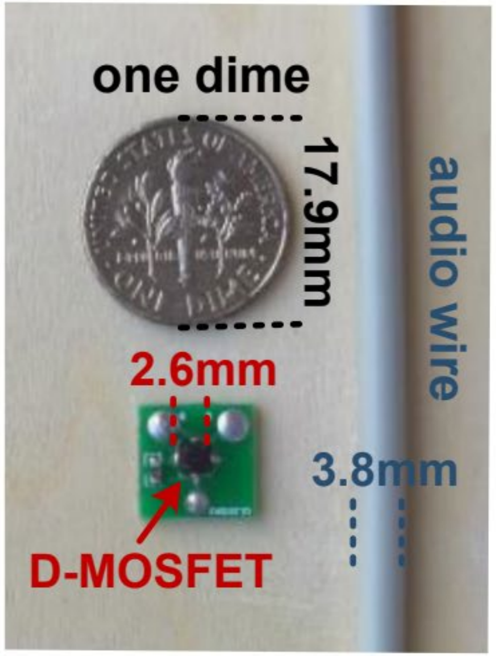
System Design

PRINCETON UNIVERSITY
山东大学 SHANDONG UNIVERSITY

Fabrication of D-MOSFET-based Retroreflector:



(a) retroreflector circuit diagram



(b) prototype

System Design

RF Signal Emission and Receiving

The dipole antenna, with a length of L, would resonate at the odd multiples of half-wavelength for the RF signal. Then, the candidate resonance frequency f_r can be expressed as follows:

$$f_r = \frac{1}{2} \cdot (2n - 1) \cdot c/L$$

We choose 2.25GHz, which resonates at 15 multiples of half-wavelength for a typical 1-m cable.

RF Signal Pre-processing

- Low-pass Filter
- Remove the DC component
- Divide the signal into segments with fixed length
- Calculate the mel-spectrogram of each segment

System Design

Rethink non-linear transfer curve before audio Reconstruction

$$y(t)=\begin{cases}\alpha\cdot V_{gs}+\beta & V_{gs}\geq 0\\ -e^{\gamma\cdot V_{gs}}\cdot V_{gs} & V_c\leq V_{gs}<0\\ 0 & V_{gs}<V_c\end{cases}$$

The negative part kept, but distorted non-linearly.

System Design

PRINCETON UNIVERSITY 山东大学 SHANDONG UNIVERSITY

Rethink non-linear transfer curve before audio Reconstruction

$$y(t) = \begin{cases} \alpha \cdot V_{gs} + \beta & V_{gs} \geq 0 \\ -e^{\gamma \cdot V_{gs}} \cdot V_{gs} & V_c \leq V_{gs} < 0 \\ 0 & V_{gs} < V_c \end{cases}$$

The negative part kept, but distorted non-linearly.

1. Decompose the negative part:

$$a_1(t) = -\, e^{\gamma \cdot V_{gs}(t)},\ a_2(t) = V_{gs}(t),\ y(t) = a_1(t) \cdot a_2(t)$$

System Design

PRINCETON UNIVERSITY 山东大学 SHANDONG UNIVERSITY

Rethink non-linear transfer curve before audio Reconstruction

$$y(t)=\begin{cases}\alpha\cdot V_{gs}+\beta & V_{gs}\geq 0\\ -e^{\gamma\cdot V_{gs}}\cdot V_{gs} & V_c\leq V_{gs}<0\\ 0 & V_{gs}<V_c\end{cases}$$

The negative part kept, but distorted non-linearly.

1. Decompose the negative part:

$$a_1(t)=-\,e^{\gamma\cdot V_{gs}(t)},\ a_2(t)=V_{gs}(t),\ y(t)=a_1(t)\cdot a_2(t)$$

2. Convolutional operation:

$$\begin{aligned}Y(f)&=\mathscr{F}\{a_1(t)\cdot a_2(t)\}\\&=A_1(f)*A_2(f)=A_1(f)*\mathscr{F}\{V_{gs}(t)\}\end{aligned}$$

The original Frequency spectrogram ($\mathscr{F}\{V_{gs}(t)\}$)

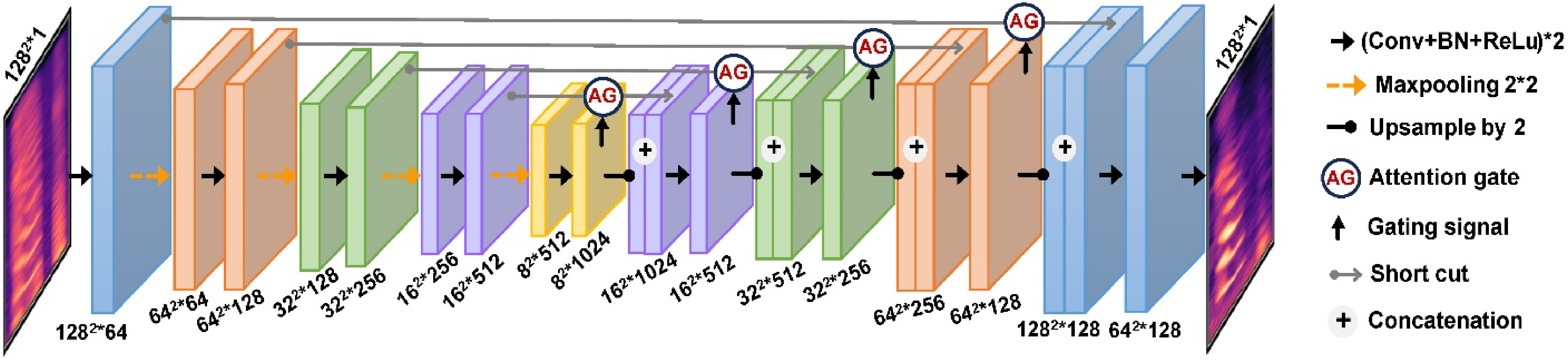
How to compute? ($A_1^{-1}(f)$)

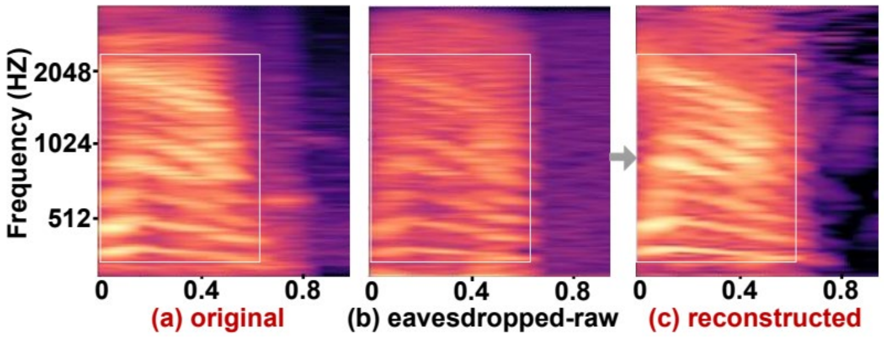
$$\begin{aligned}A_1^{-1}(f)*Y(f)&=A_1^{-1}(f)*A_1(f)*\mathscr{F}\{V_{gs}(t)\}\\&=\delta(f)*\mathscr{F}\{V_{gs}(t)\}=\mathscr{F}\{V_{gs}(t)\}\end{aligned}$$

System Design

PRINCETON UNIVERSITY 山东大学 SHANDONG UNIVERSITY

Audio Reconstruction

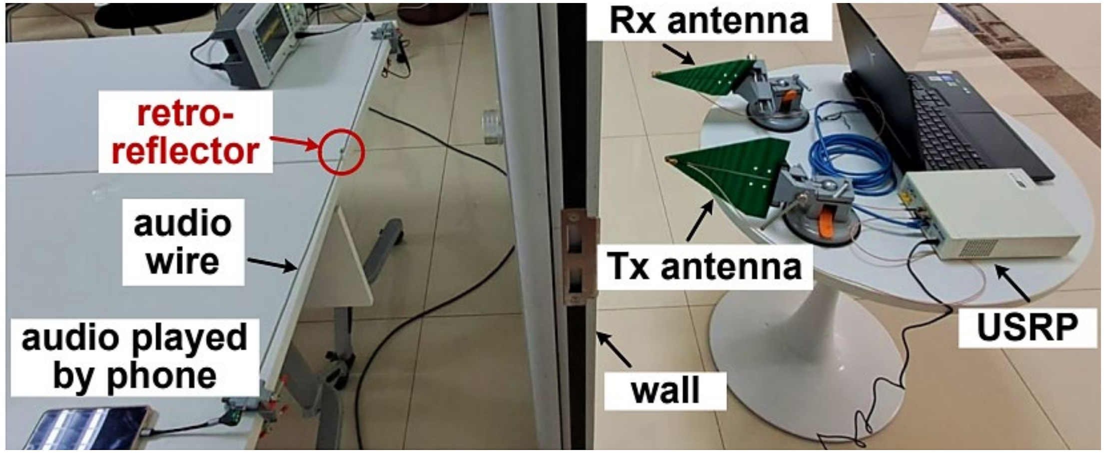




Griffin-Lim algorithm

Experiment Setup

- Audio dataset
  - Free Spoken Digit Dataset (FSDD)
  - Speech Commands Dataset (SCD)
- Metrics:
  - Mel-cepstral distortion (MCD): A lower value indicates a better reconstruction performance.
  - Mean opinion score (MOS): A higher value indicates a better reconstruction performance.
  - Signal-to-noise ratio (SNR): A higher value indicates a better reconstruction performance.
  - Peak signal-to-noise ratio (PSNR): A higher value indicates a better reconstruction performance.
  - Accuracy and F1-score: A higher value indicates a better speech command classification performance.



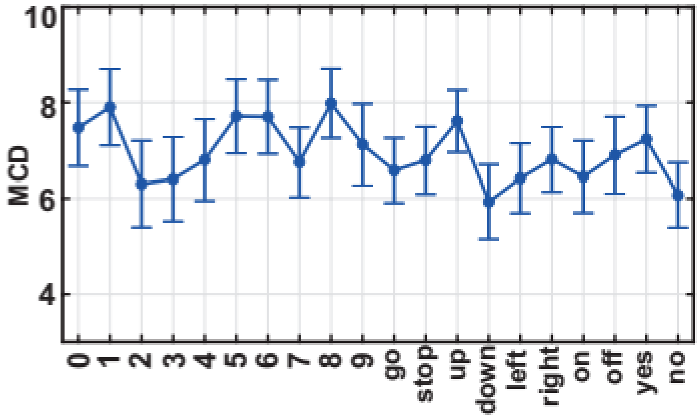
Experiment

Comparison between E and D MOSFETs:

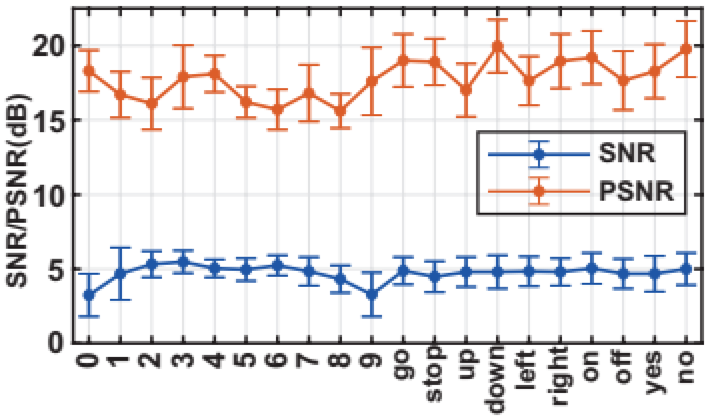
COMPARISON OF MCD USING DIFFERENT MOSFETS

MOSFET type	E-MOSFET	D-MOSFET	
		Raw	**Reconstructed**
MCD	23.1	17.0	**6.8**

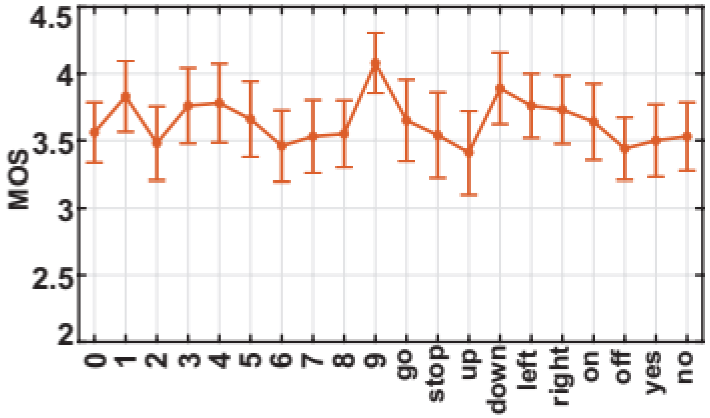
Speech reconstruction performance:



MCD for each speech command



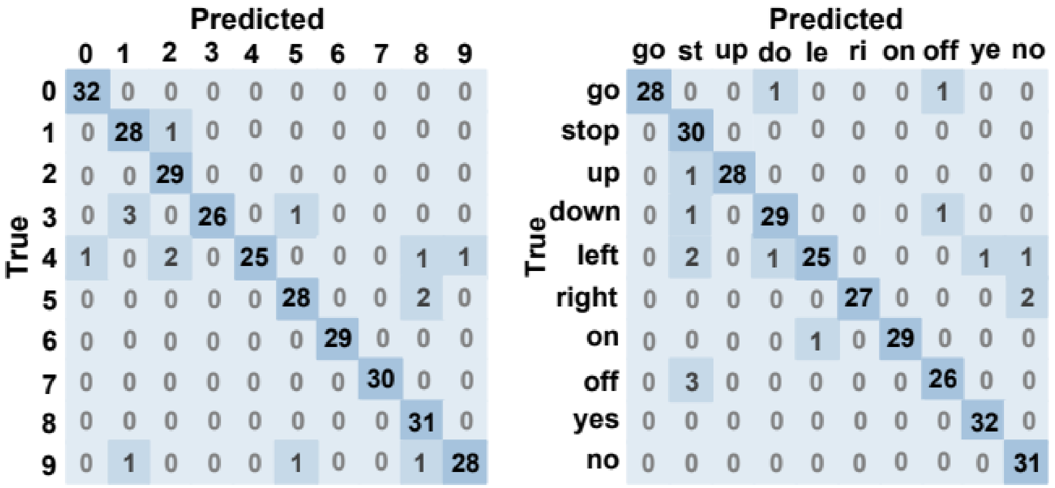
SNR/PSNR for each speech command



MOS for each speech command

Experiment

Speech command classification performance:



(a) 10 digit commands (b) 10 action commands

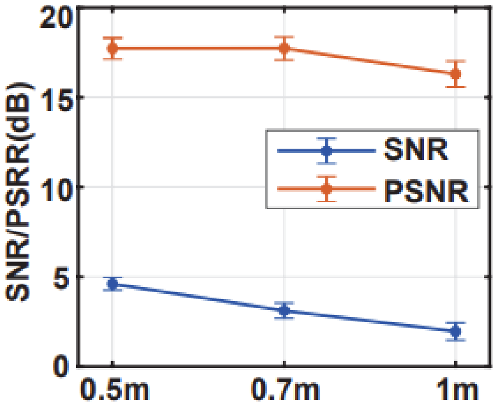
Confusion matrix of (a) digit (b) action commands

F1-SCORE OF SPEECH COMMAND RECOGNITION

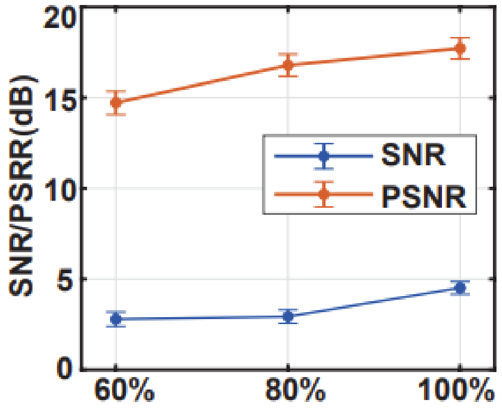
Command	0	1	2	3	4	5	6	7	8	9
F1-score	0.98	0.92	0.95	0.93	0.91	0.93	1.0	1.0	0.94	0.93
Command	**go**	**stop**	**up**	**down**	**left**	**right**	**on**	**off**	**yes**	**no**
F1-score	0.97	0.91	0.98	0.94	0.91	0.96	0.95	0.91	0.98	0.95

F1-score of digit and action commands

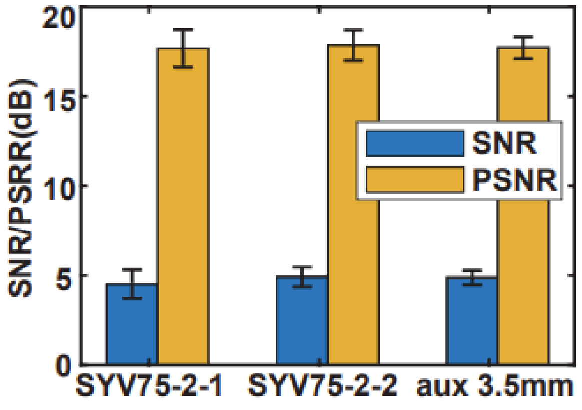
Impact of practical factors:



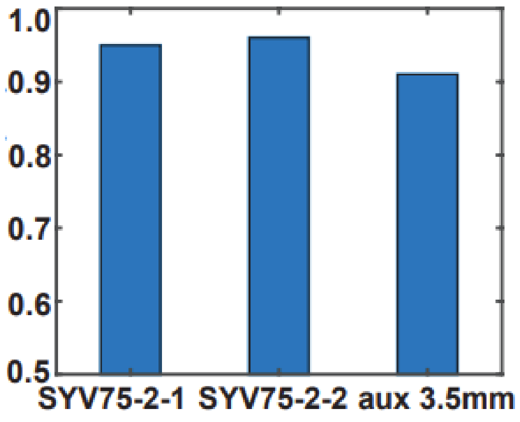
Distance



Volume



Wire type



Recognition accuracy of wire types

Demo

Speech Command	Original	Eavesdropped	Reconstructed
'Left'			
'Off'			
'Yes'			

Conclusion

- We propose the first wired audio eavesdropping attack, RF-Parrot, with a simple yet effective D-MOSFET retroreflector.
- RF-Parrot achieve 95% accuracy in identifying speech commands, by leveraging the encoder-decoder neural network with convolutional layers.
- We believe this work will raise awareness of the potential safety hazards of earphone systems.

Thanks for your listening!

Q&A

Yanni Yang, **Genglin Wang**, Zhenlin An, Guoming Zhang, Xiuzhen Cheng, Pengfei Hu