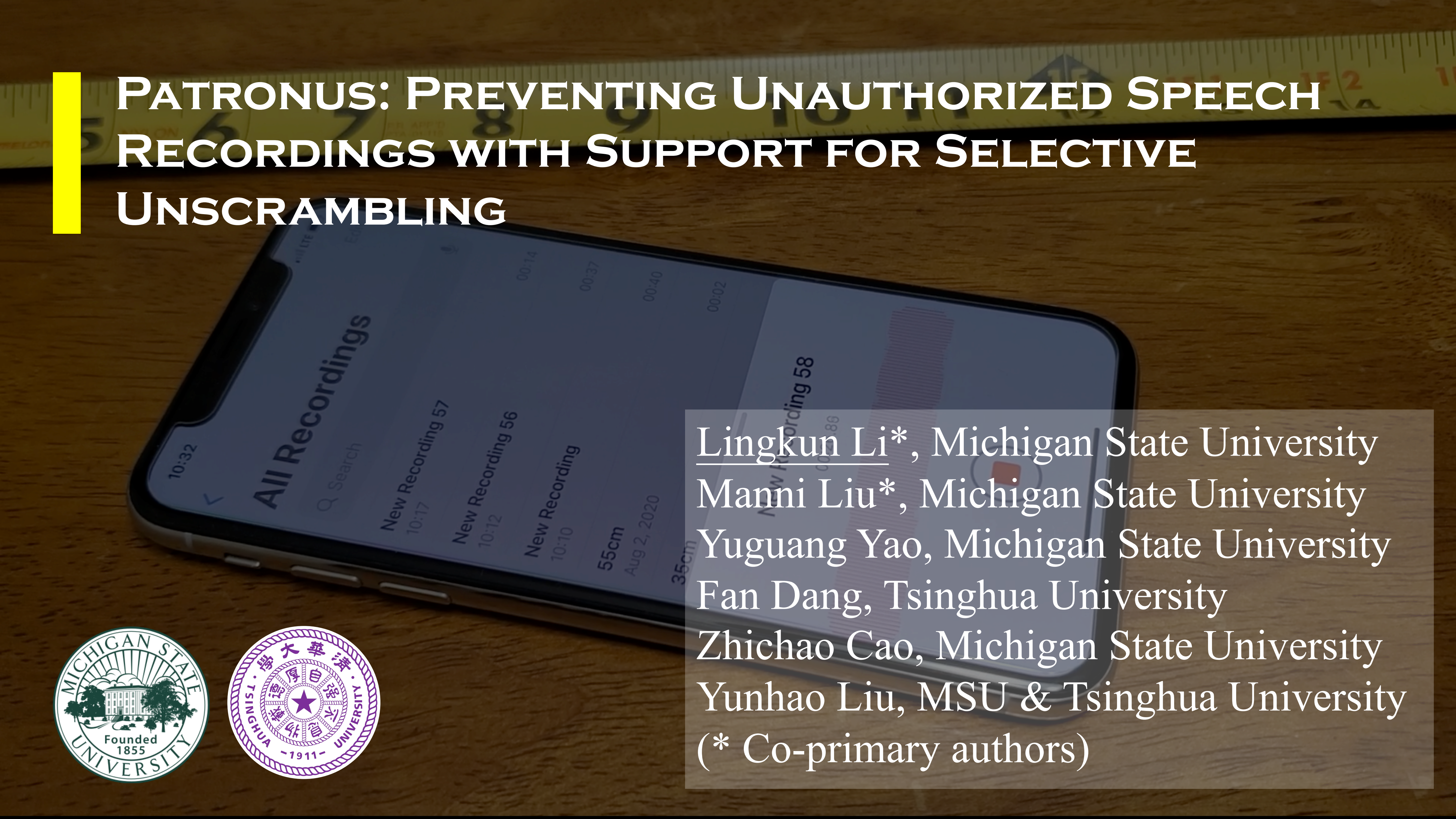
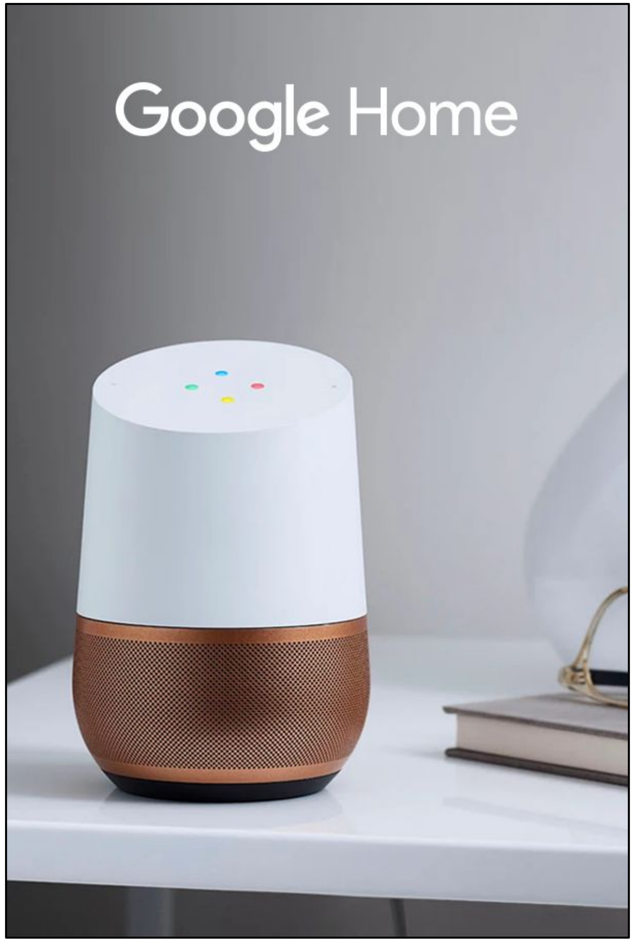
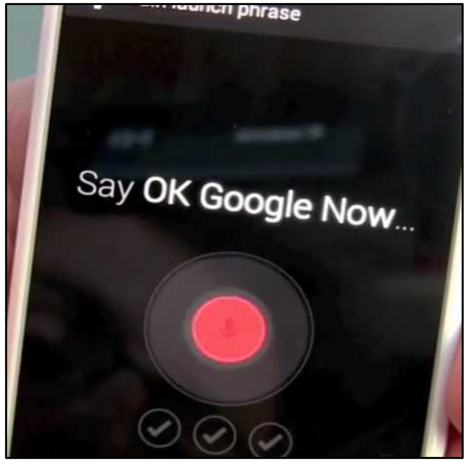
Patronus: Preventing Unauthorized Speech Recordings with Support for Selective Unscrambling



Lingkun Li*, Michigan State University
Manni Liu*, Michigan State University
Yuguang Yao, Michigan State University
Fan Dang, Tsinghua University
Zhichao Cao, Michigan State University
Yunhao Liu, MSU & Tsinghua University
(* Co-primary authors)

Smart devices lead us to privacy risks





Anti-secret Audio Recording

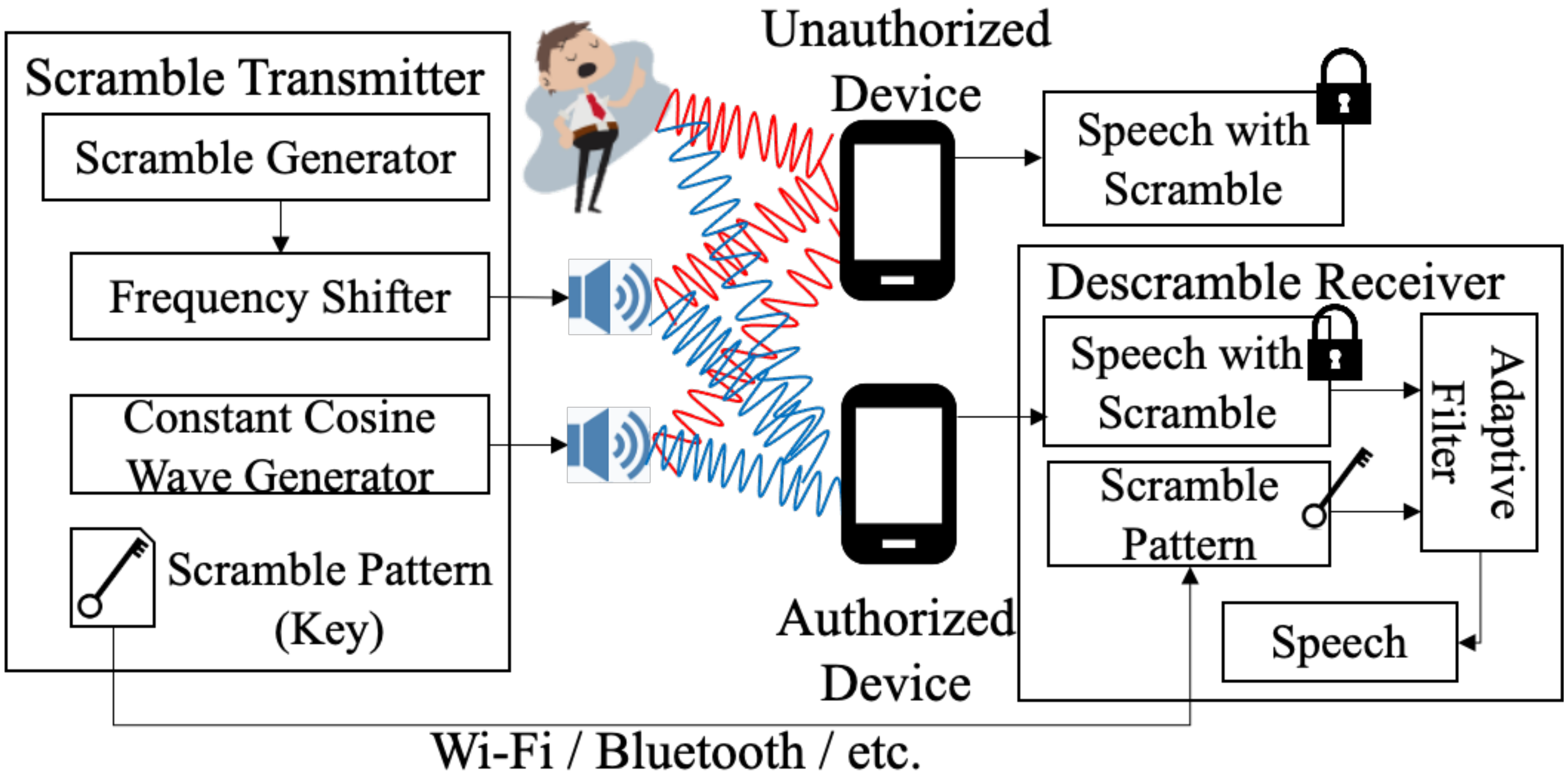
Requirements:

(1) Normal human conversation should be unaffected.

(2) Unauthorized devices should not be able to make a clear recording.

(3) Authorized devices should be able to make a clear recording of any conversation protected by anti-recording solution.

Patronus


Scramble Transmitter
Scramble Generator
Frequency Shifter
Constant Cosine Wave Generator
Scramble Pattern (Key)
Unauthorized Device
Speech with Scramble
Descramble Receiver
Speech with Scramble
Scramble Pattern
Adaptive Filter
Speech
Authorized Device
Wi-Fi / Bluetooth / etc.

Scrambling: Nonlinear effect

Input signal: $s(t) = \cos(2\pi f_1 t) + \cos(2\pi f_2 t)$

Output signal: $\mathrm{y}(t) \approx A_2 \cos[2\pi(f_1 - f_2)t]$

This study: $\mathrm{y}(t) \approx \sum_{j=1}^{n} A_j \cos[2\pi j(f_1 - f_2)t]$

It provide us with the possibility of using ultrasound to scramble COTS microphones.

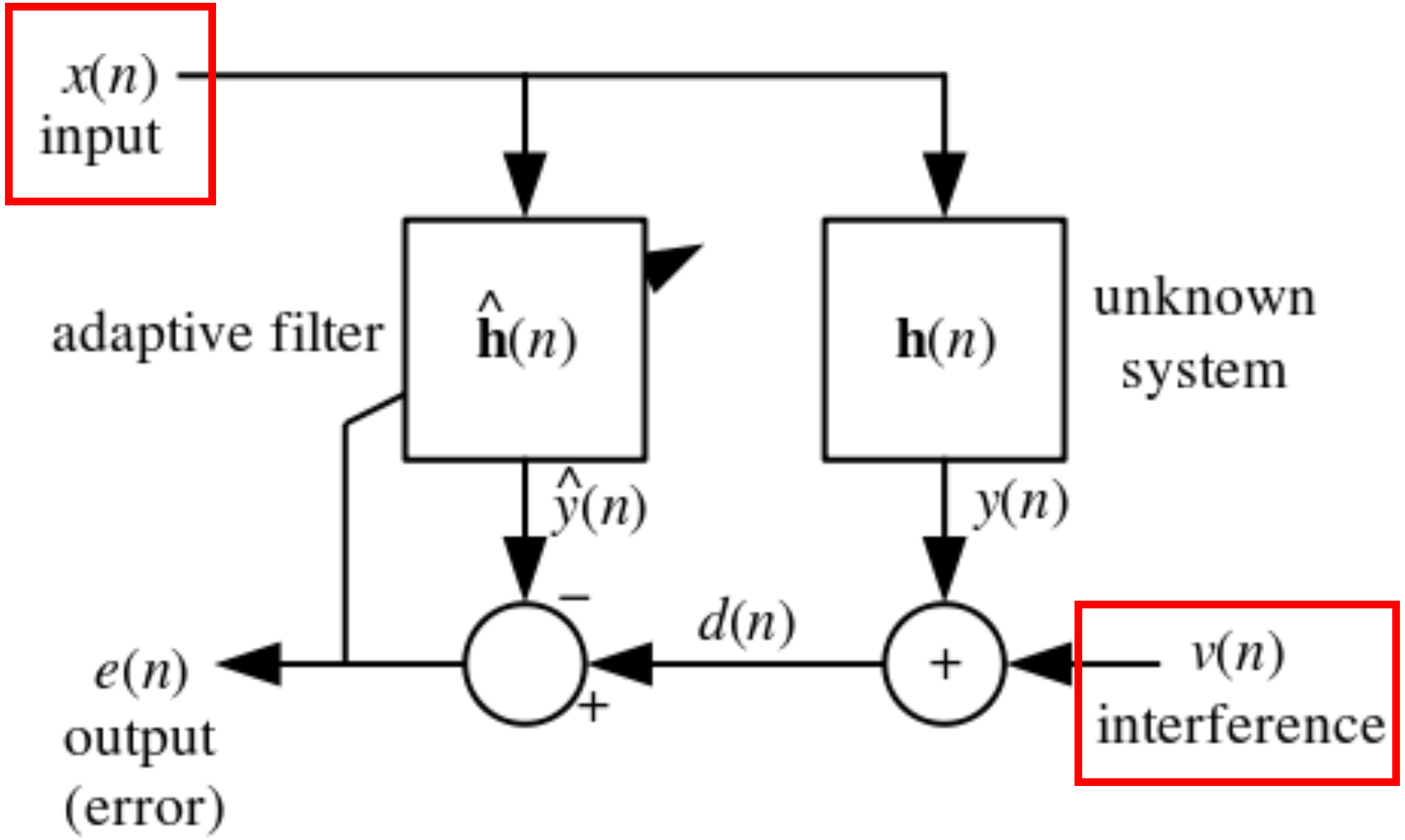
Related works:

- BackDoor (MobiSys 2017)
- LipRead (NSDI 2018)
- DolphinAttack (CCS 2018)
- UPS+ (MobiCom 2019)
- SurfingAttack (NDSS 2020)

Decrambling

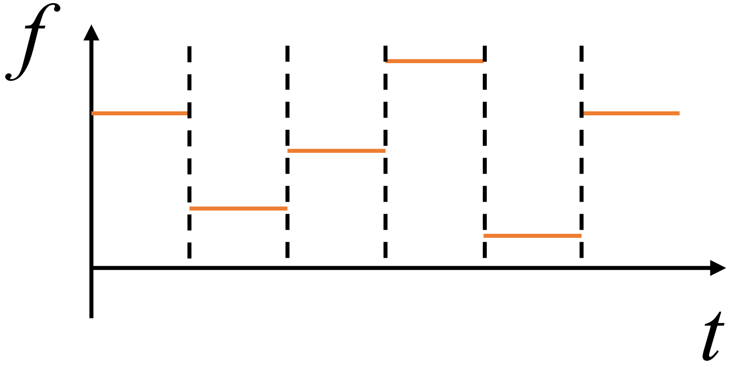
Normalized Least Mean Square (NLMS) filter

Scrambled recording

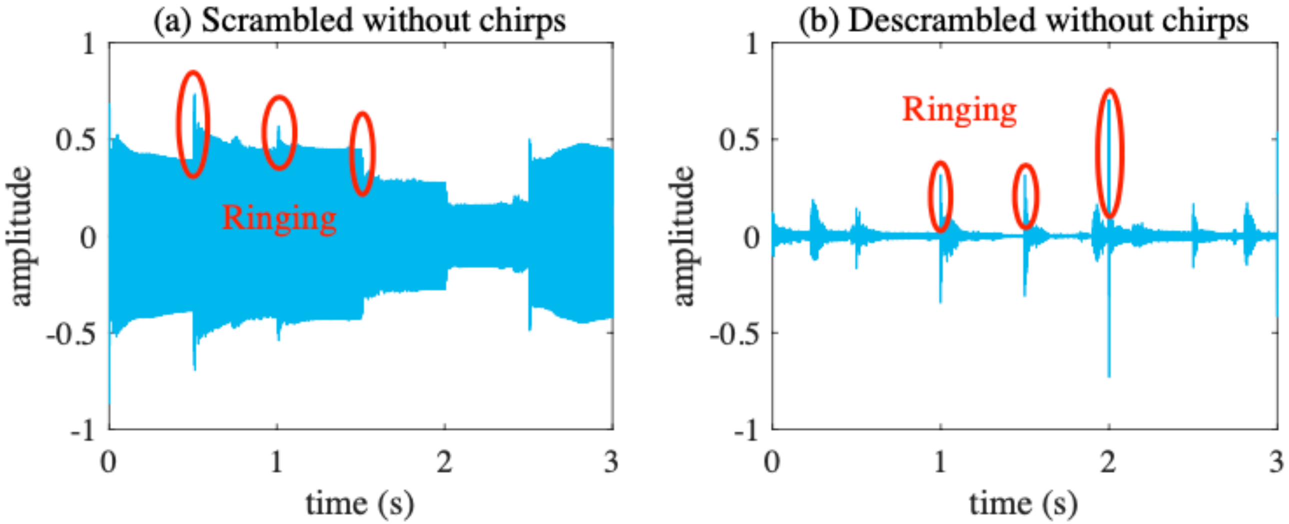


Scramble wave

Challenge 1: Ringing Effect

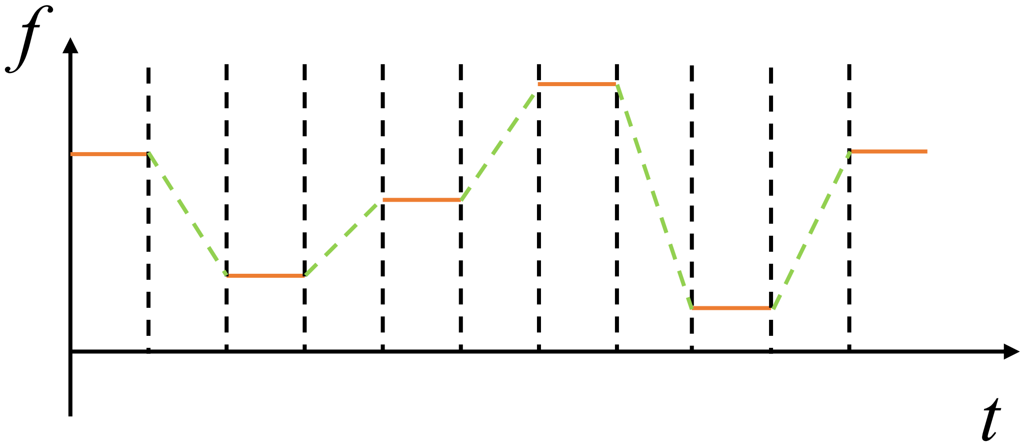


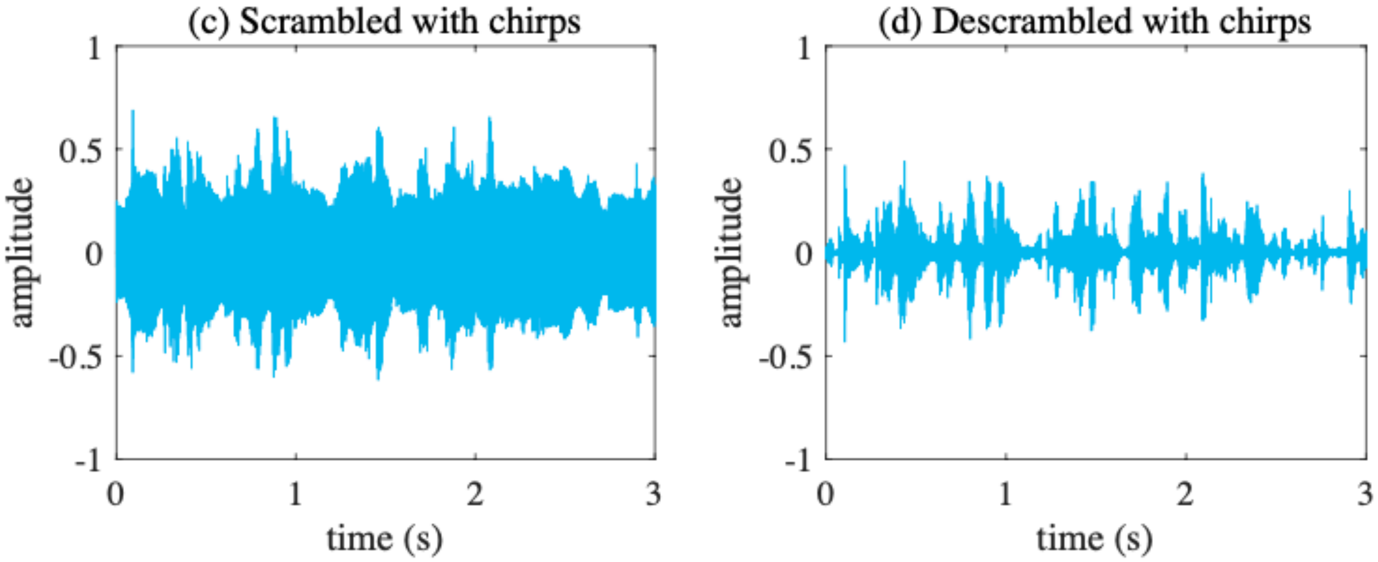
Scramble pattern



Recording with ringing

Challenge 1: Ringing Effect





Chirp-smoothed scramble

Recording without ringing

Challenge 2: STFT Attack

First step:

Apply Short Time Fourier Transform to the recording.

Analyze the scramble waveform.

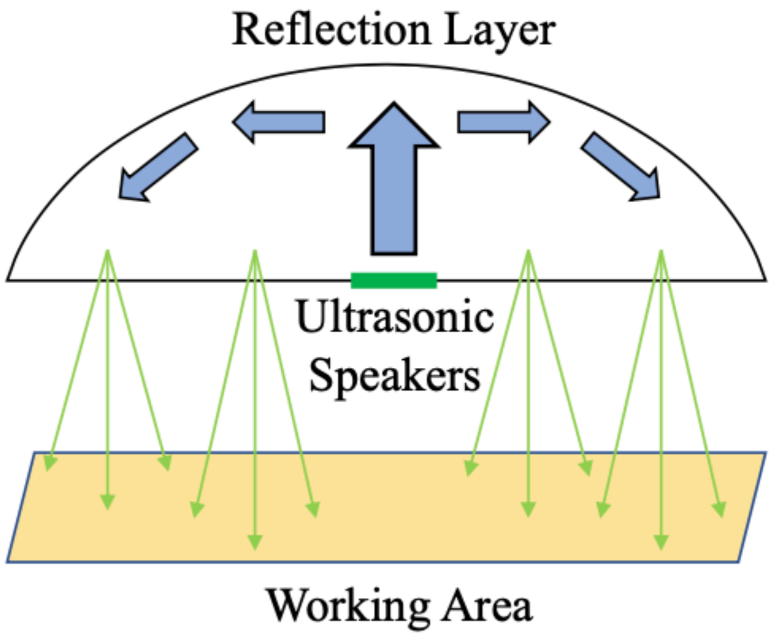
Second step:

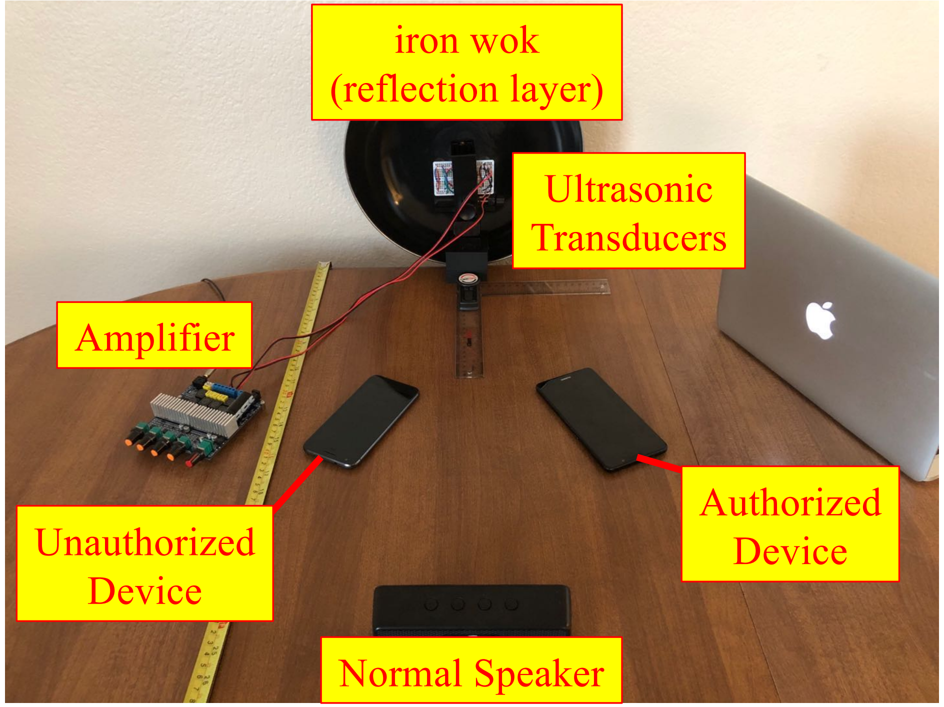
Reconstruct the scramble and use NLMS filter to cancel it out.

Solution:

Fine-tune the duration of each frequency component.

Challenge 3: Limited Working Area





Experiment Settings

Human speech material: 55 news segments, each 30 seconds long.

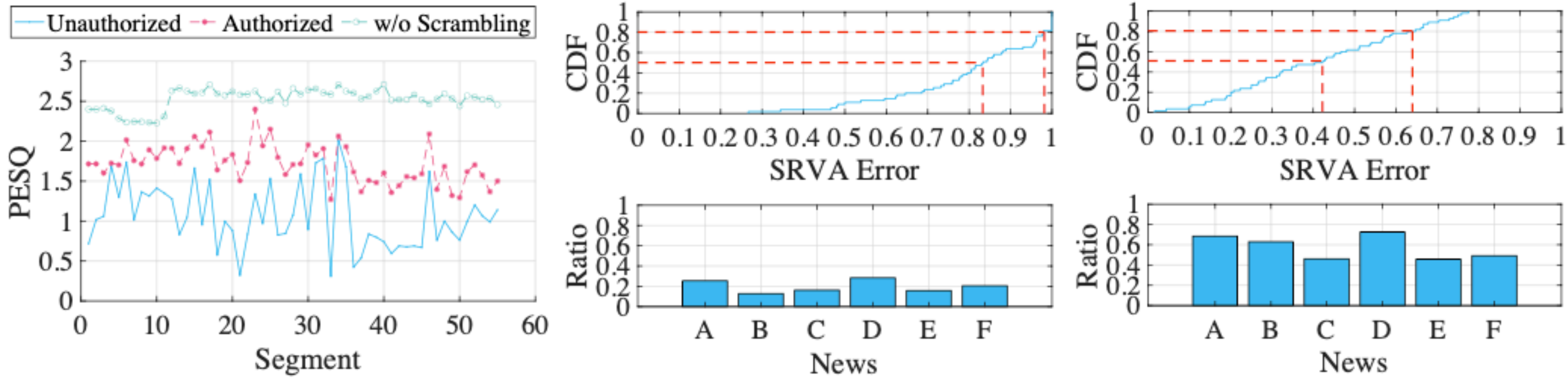
Human speech envrionment simulation:

- Speaker
- human reading

Two metrics:

- Perceptual Evaluation of Speech Quality (PESQ)
- Speech Recognition Vocabulary Accuracy (SRVA)

Speech played by speaker



(a) PESQ

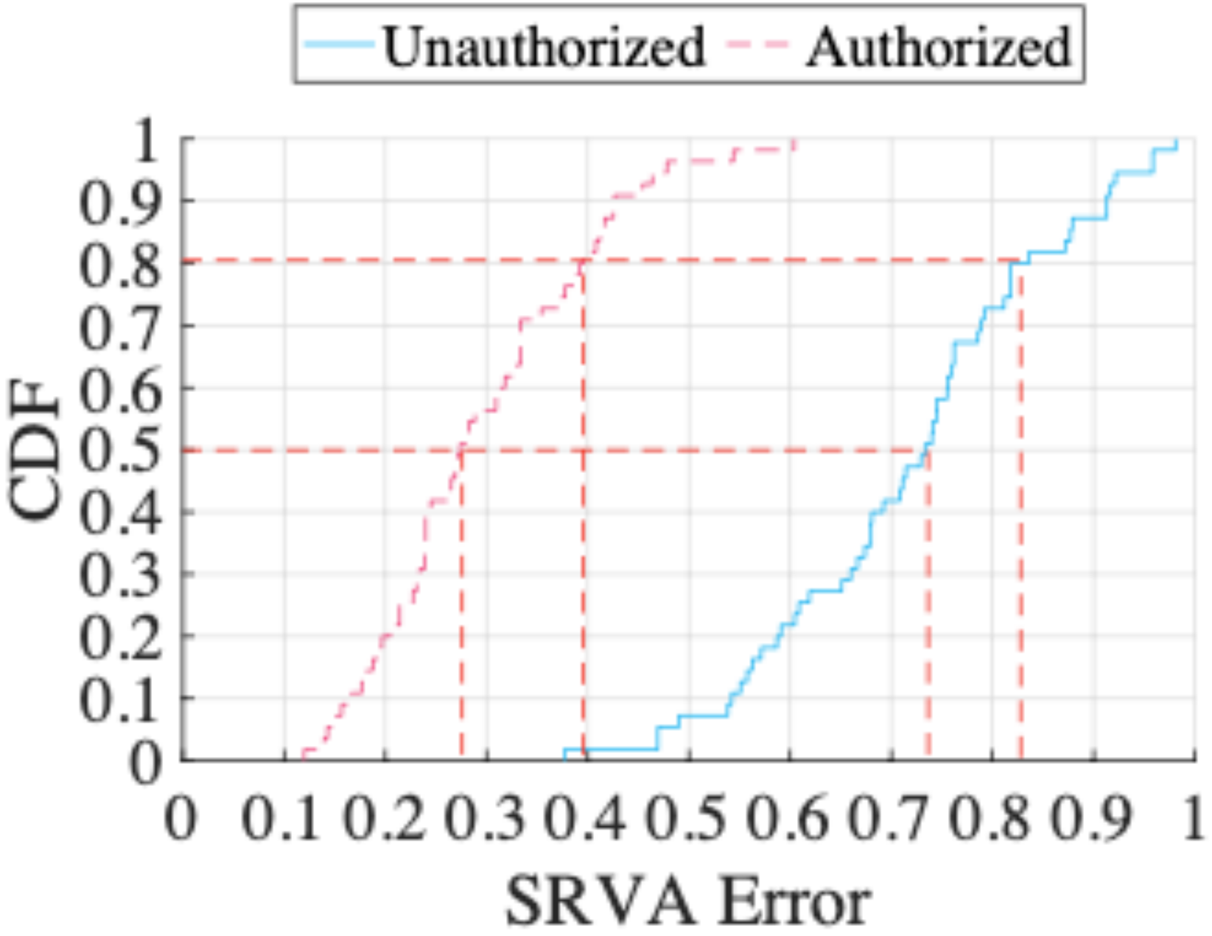
(b) Scrambled

(c) Descrambled

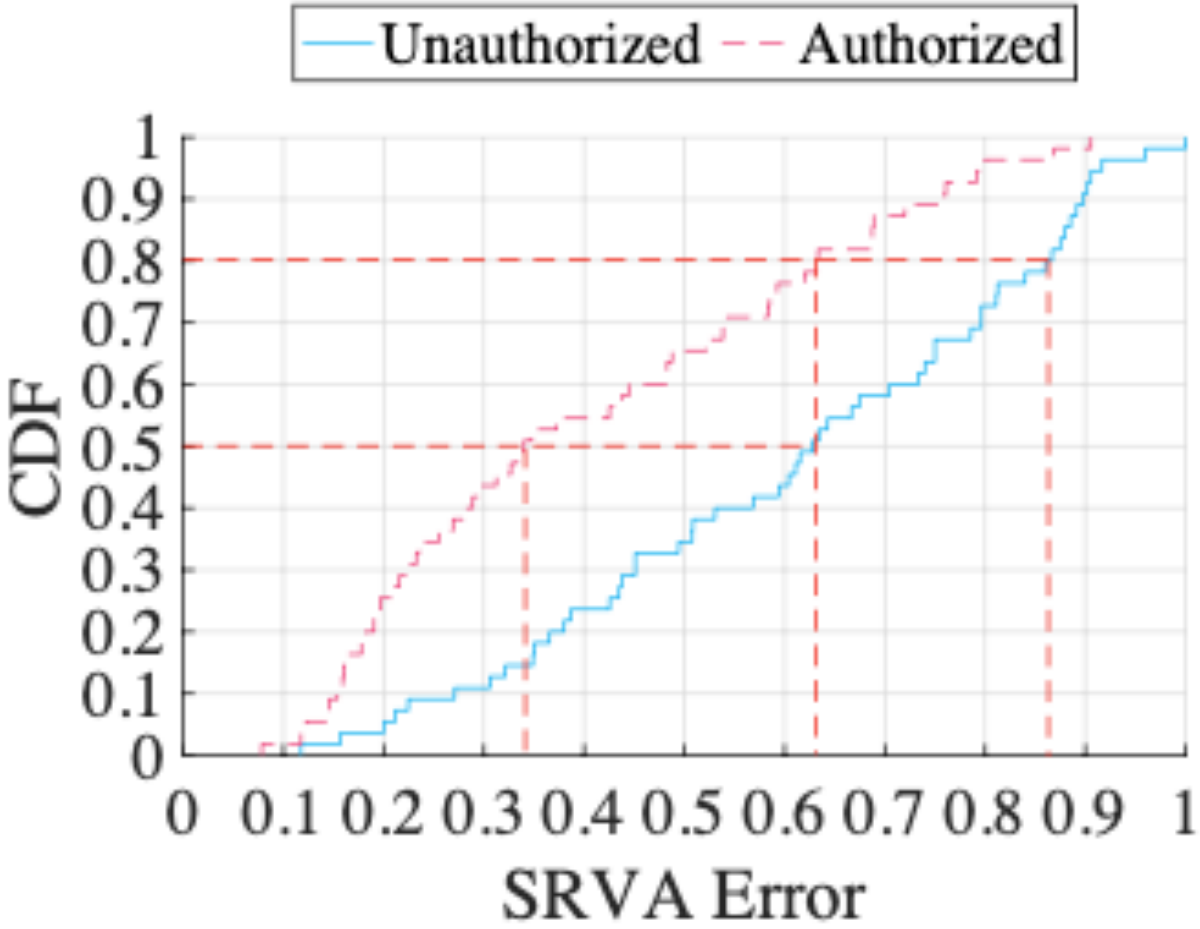
PESQ: 80% lower than 1.5 for unauthorized devices, only 16.3% lower than 1.5 for authorized devices.

SRVA: Overall, every recorded news with the authorized device have at least 2x of SRVA to the unauthorized device.

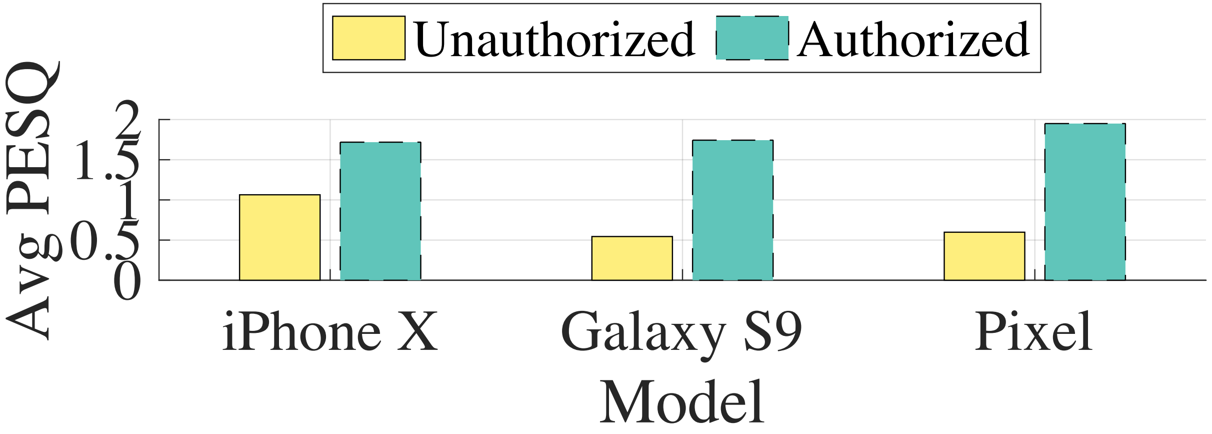
Speech read by human

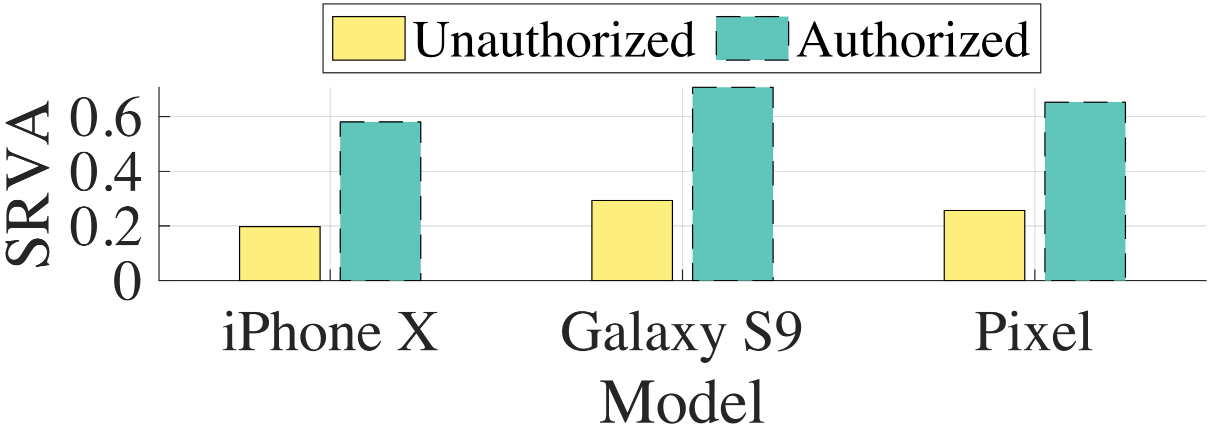


Human Recognition



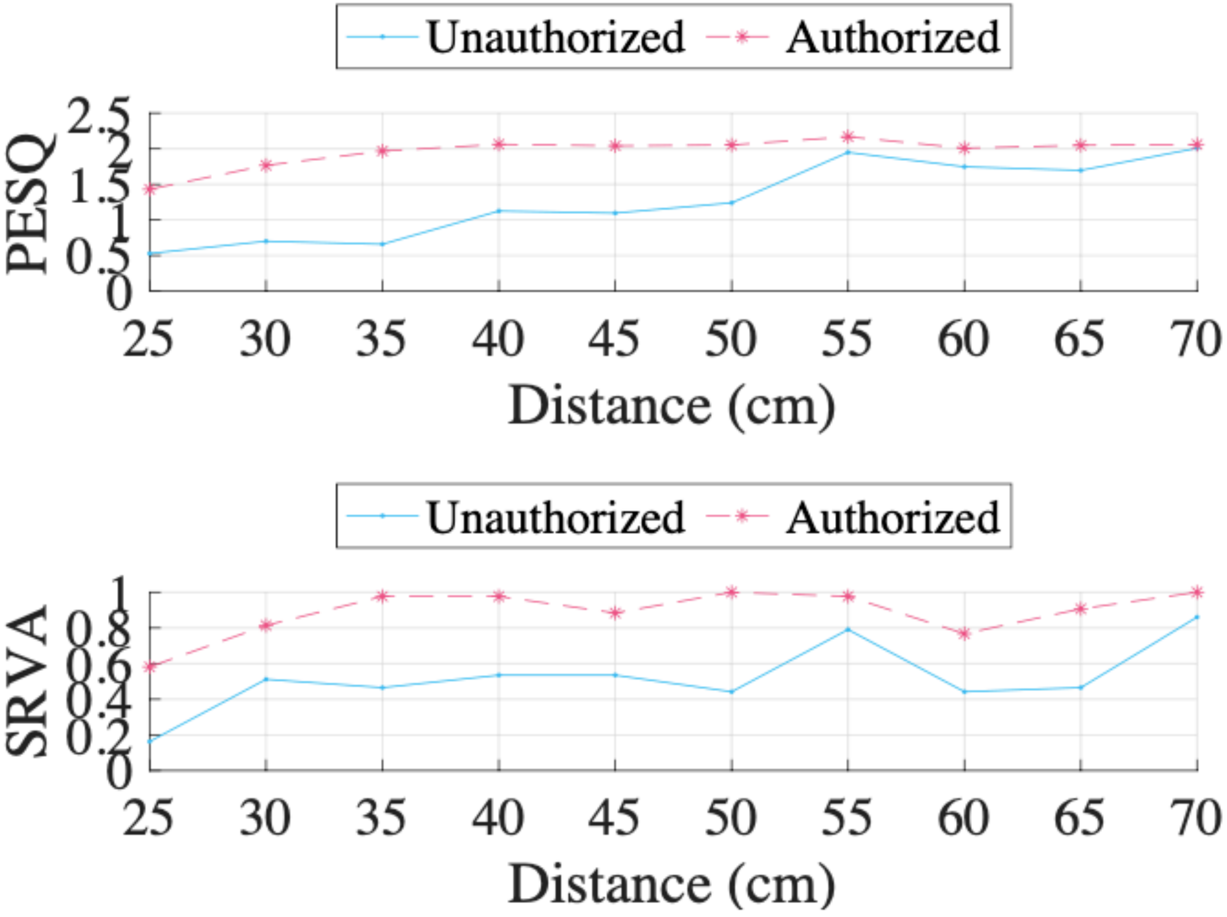
Different models



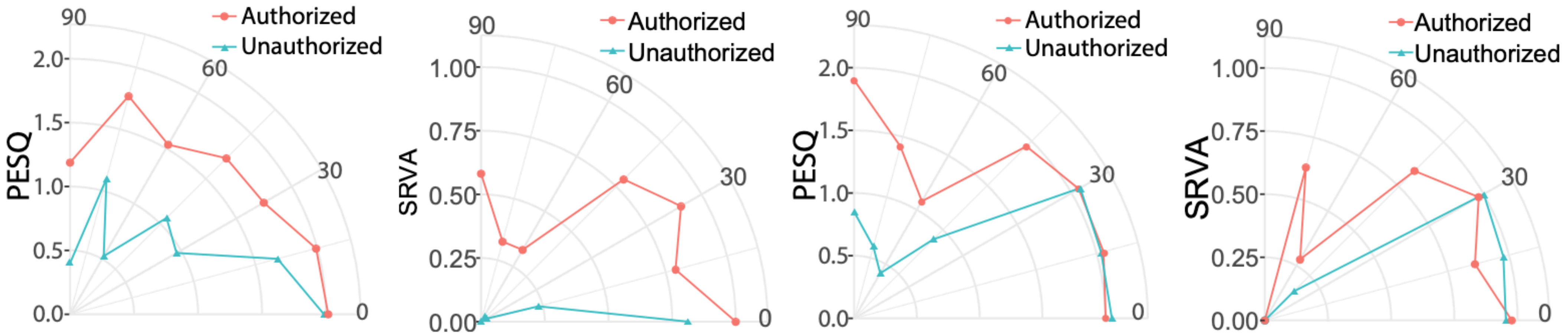


Patronus is stable among different models.

Different distances



Working area

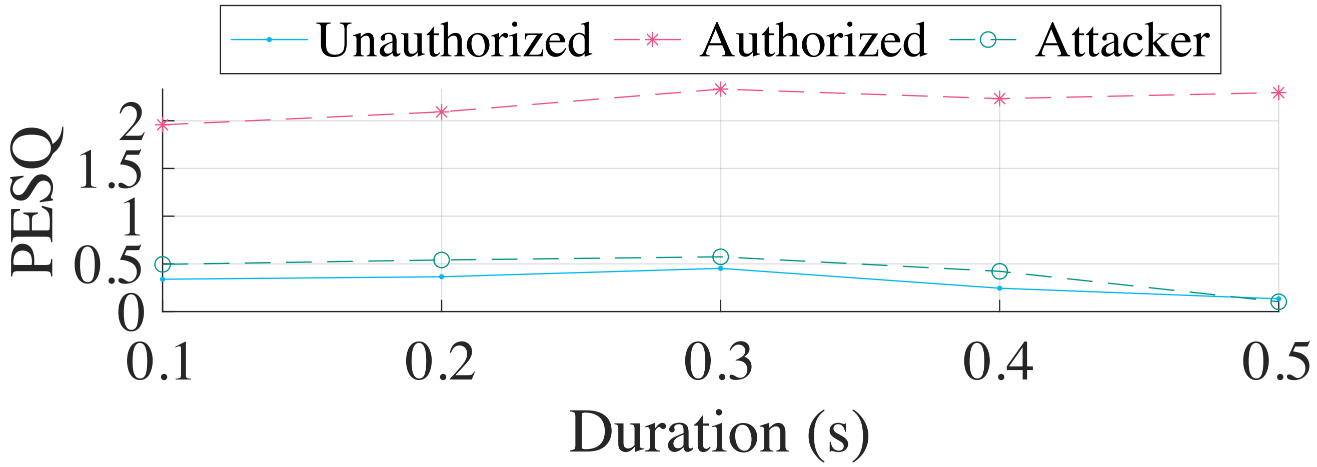


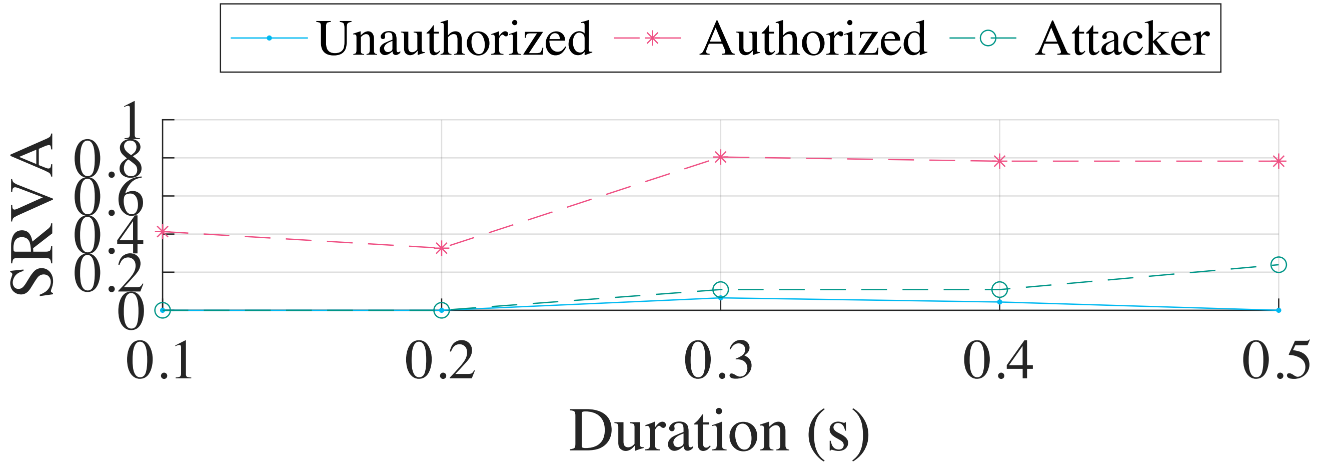
(a) PESQ (b) SRVA (c) PESQ (d) SRVA

covered

uncovered

Frequency durations





Conclusion

- We do a thorough study around the nonlinear effect of ultrasound on commercial microphones.
- Based on the study, we propose an optimized configuration to generate the scramble. It would provide privacy protection against unauthorized recordings that does not disturb normal conversation.
- We use NLMS filter to cancel out the scramble for authorized devices and fine-tune the frequency duration to prevent STFT attack.
- We design a low-cost reflection layer to enlarge the working area.

Thank You

Lingkun Li*

Dept. of Computer Science and Engineering
Michigan State University
https://www.cse.msu.edu/~lilingk1

Manni Liu*, Michigan State University
liumanni@msu.edu

Yuguang Yao, Michigan State University
yaoyugua@msu.edu

Fan Dang, Tsinghua University
dangf09@gmail.com

Zhichao Cao, Michigan State University
caozc@msu.edu

Yunhao Liu, MSU & Tsinghua University
yunhaoliu@gmail.com