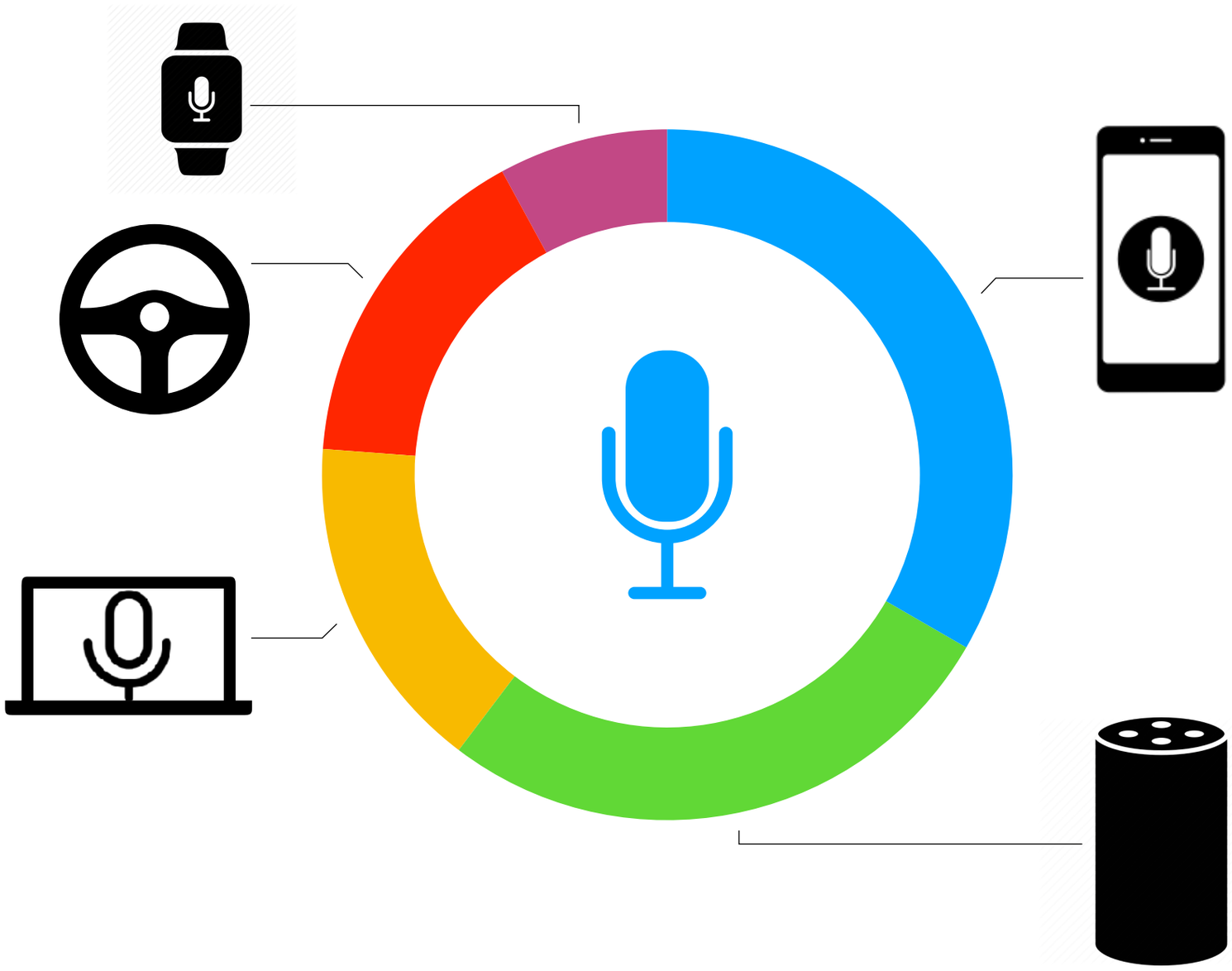
Metamorph:
Injecting Inaudible Commands into Over-the-air Voice Controlled Systems

Tao Chen[1] Longfei Shangguan[2] Zhenjiang Li[1] Kyle Jamieson[3]

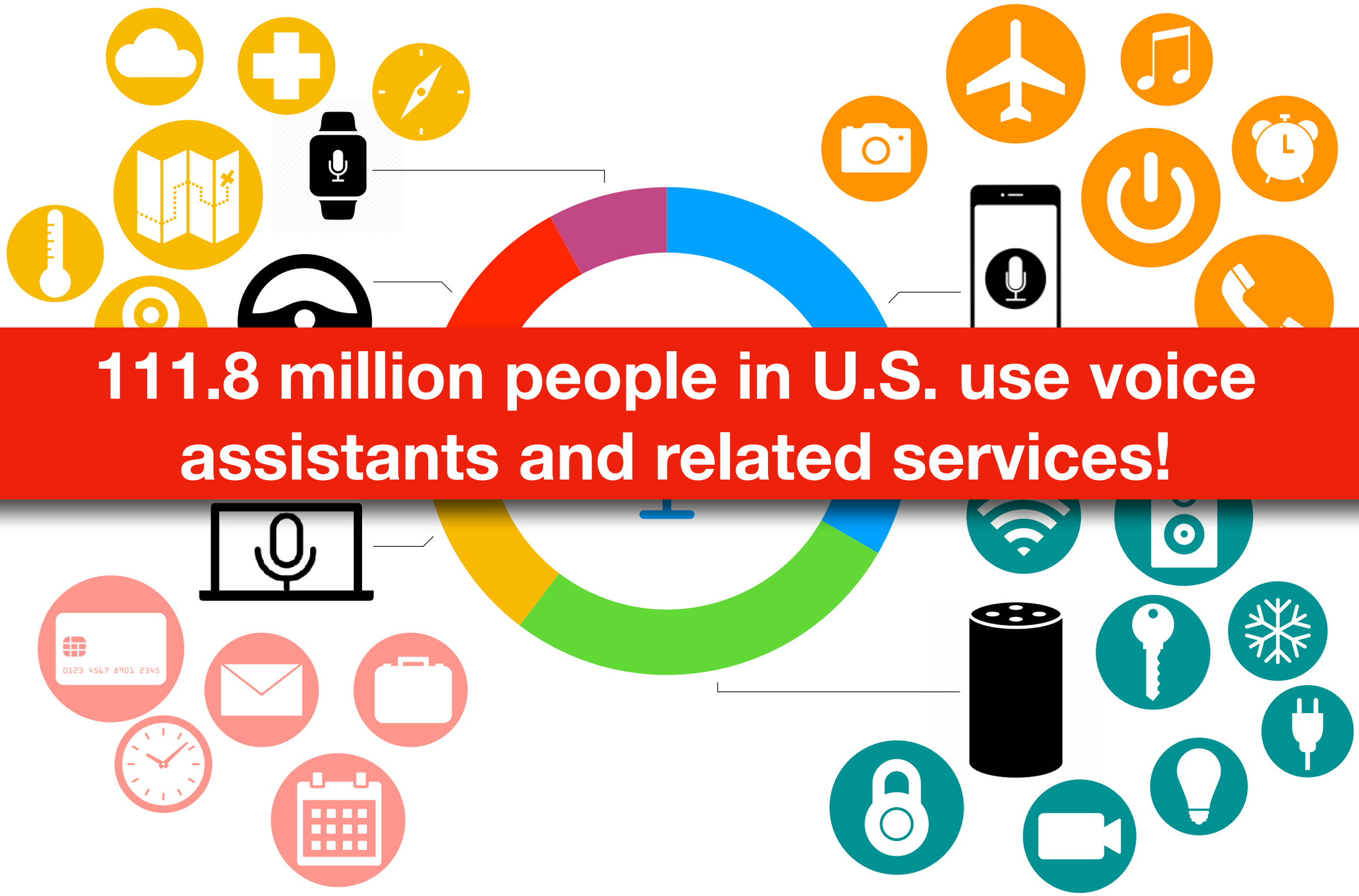
[1]City University of Hong Kong, [2]Microsoft, [3]Princeton University

Voice Assistants in Smart Home



Voice Assistants in Smart Home

Voice Assistants in Smart Home



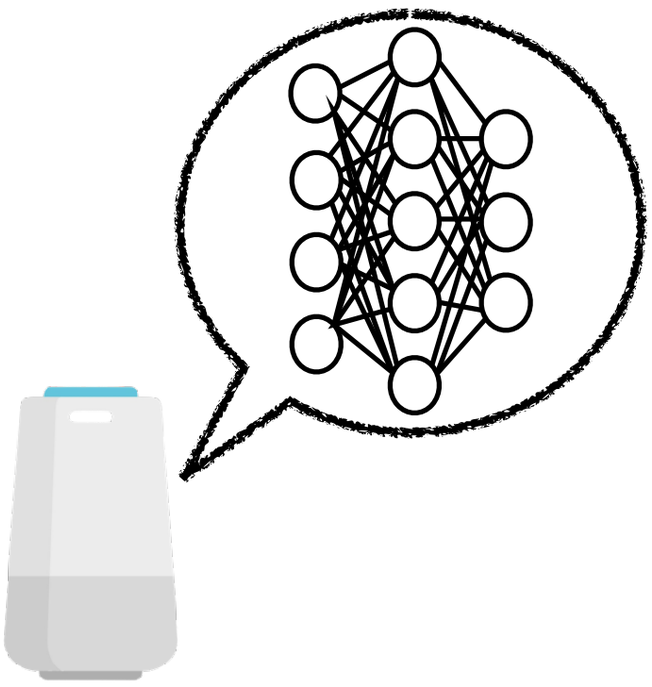
111.8 million people in U.S. use voice assistants and related services!

https://www.emarketer.com/content/us-voice-assistant-users-2019

Are they safe enough?

How to attack the voice assistant?

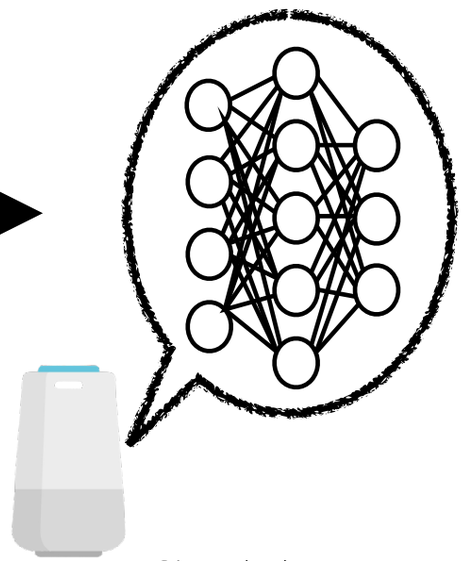
Neural networks



Speech Recognition Models (SR)

How to attack the voice assistant?

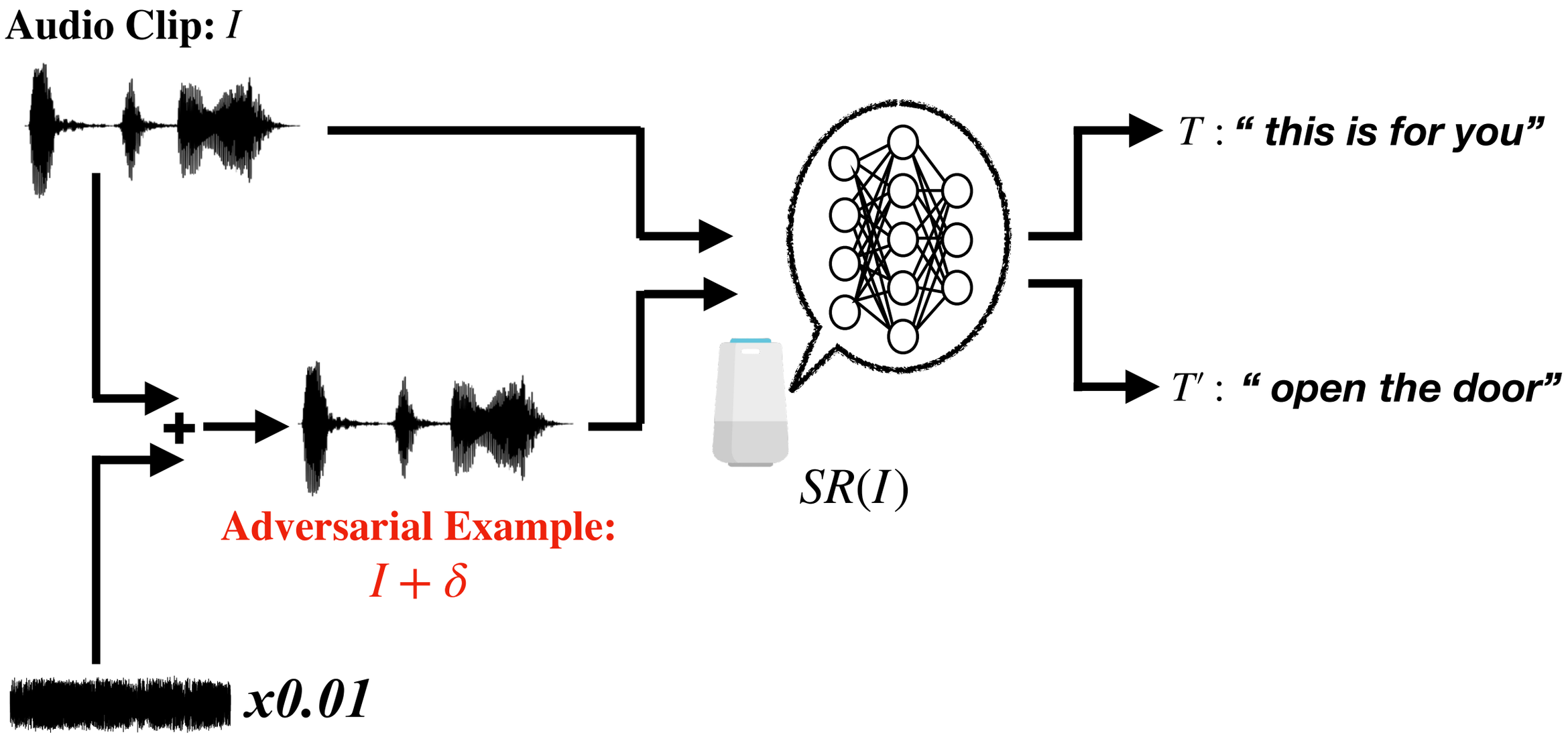
Audio Clip: I



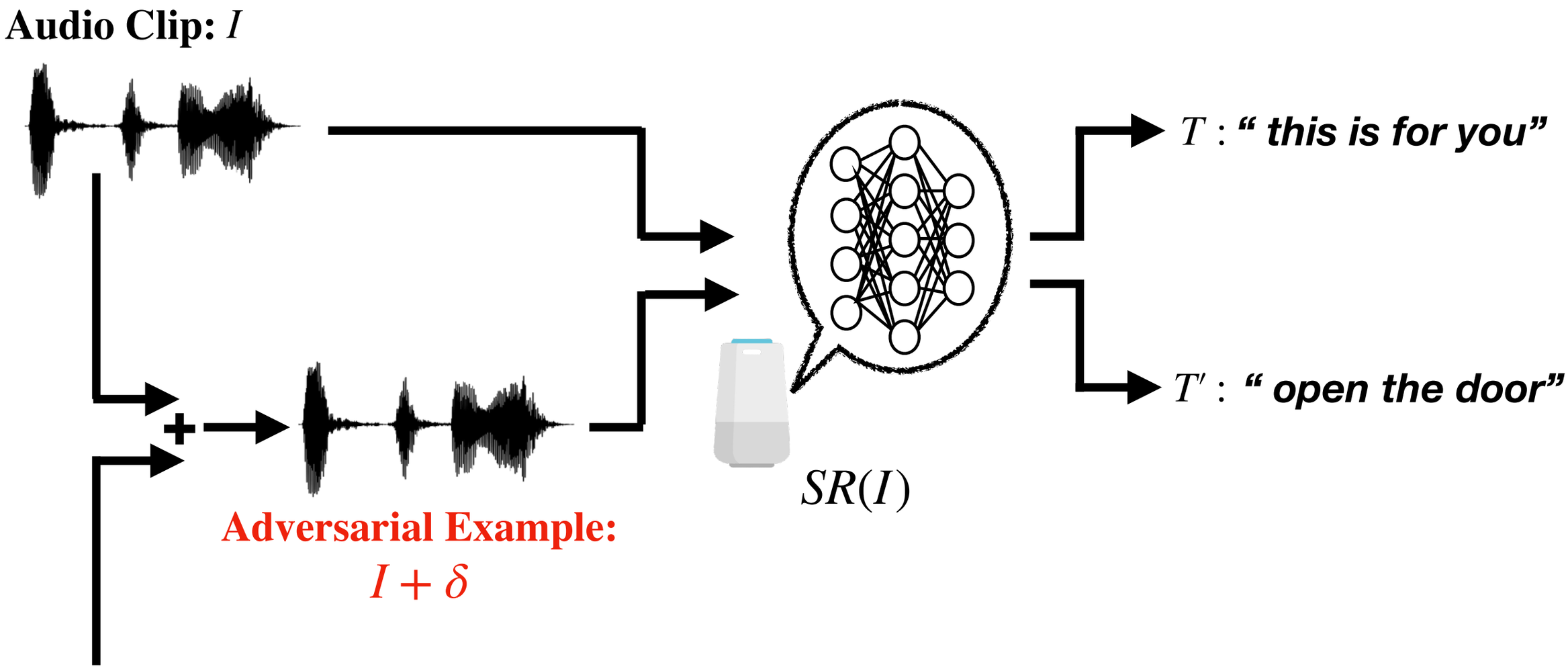
T : ***“ this is for you”***

$SR(I)$

How to attack the voice assistant?



How to attack the voice assistant?



$$\begin{aligned} \textit{minimize} \quad & dB_I(\delta), \\ \textit{such that} \quad & SR(I) = T, \\ & SR(I + \delta) = T' \end{aligned}$$

Nicholas Carlini et al. Audio Adversarial Examples, Deep Learning and Security Workshop, 2018

How to attack the voice assistant?

Audio Adversarial Attack

Nicholas Carlini et al. Audio Adversarial Examples, Deep Learning and Security Workshop, 2018

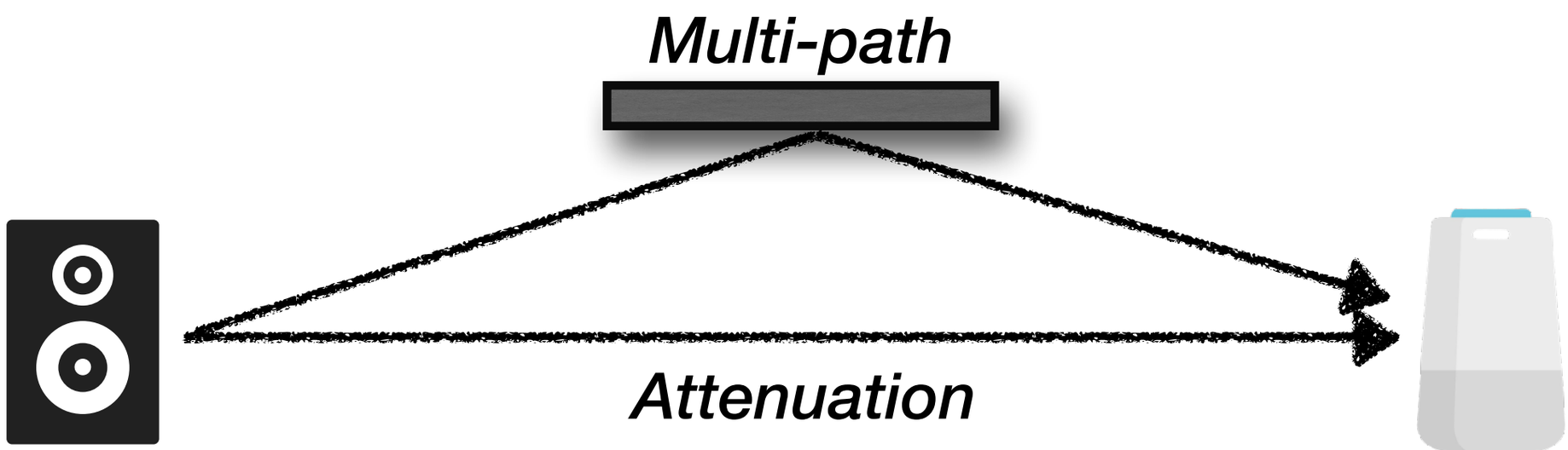
Is it a real threat? Yes!


Adversarial Example

But, failed Over-the-air!

Challenge

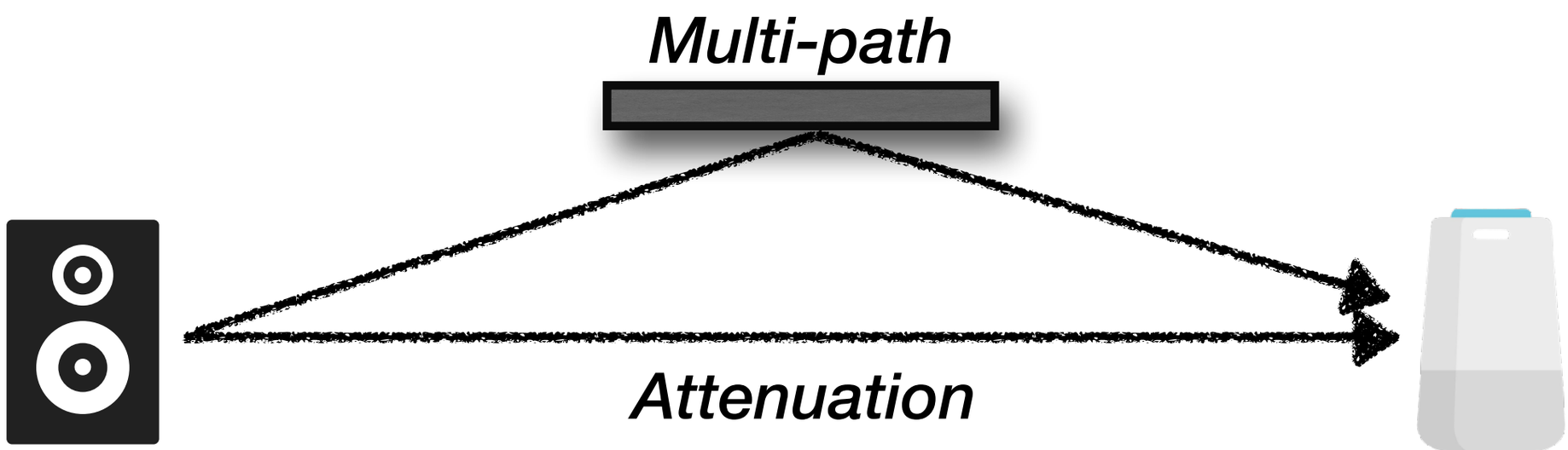
Channel Effect



Hardware Heterogeneity

Challenge

Channel Effect



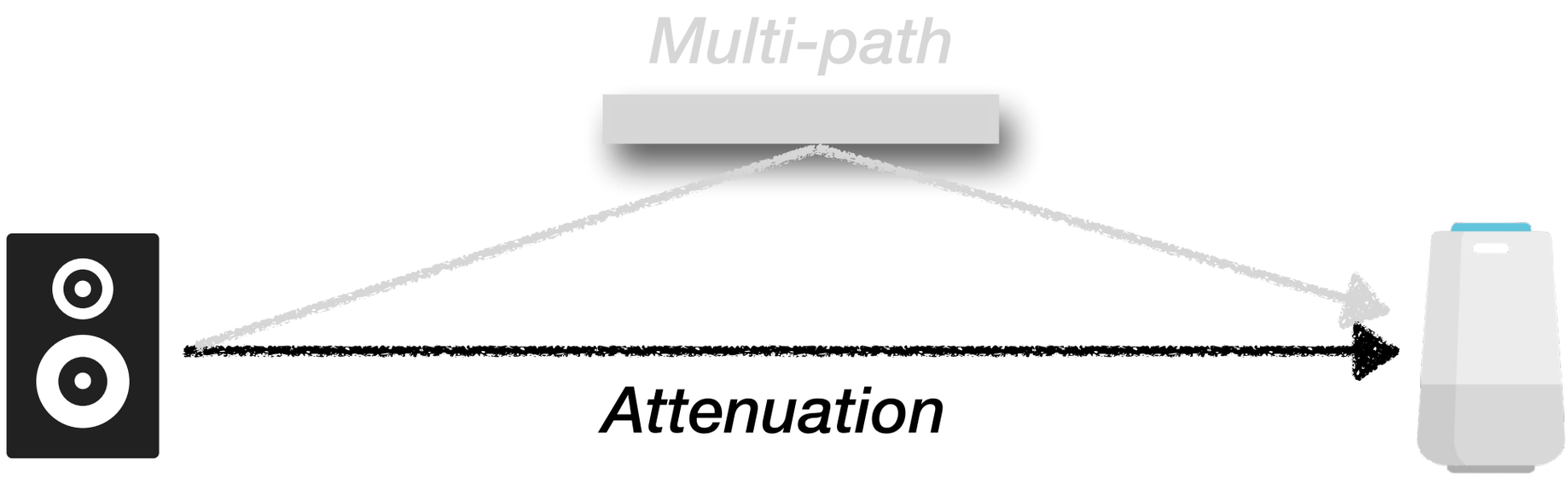
Hardware Heterogeneity

$$SR(I+\delta) \quad \textbf{vs} \quad SR(H(I+\delta))$$

H is unknown in advance!

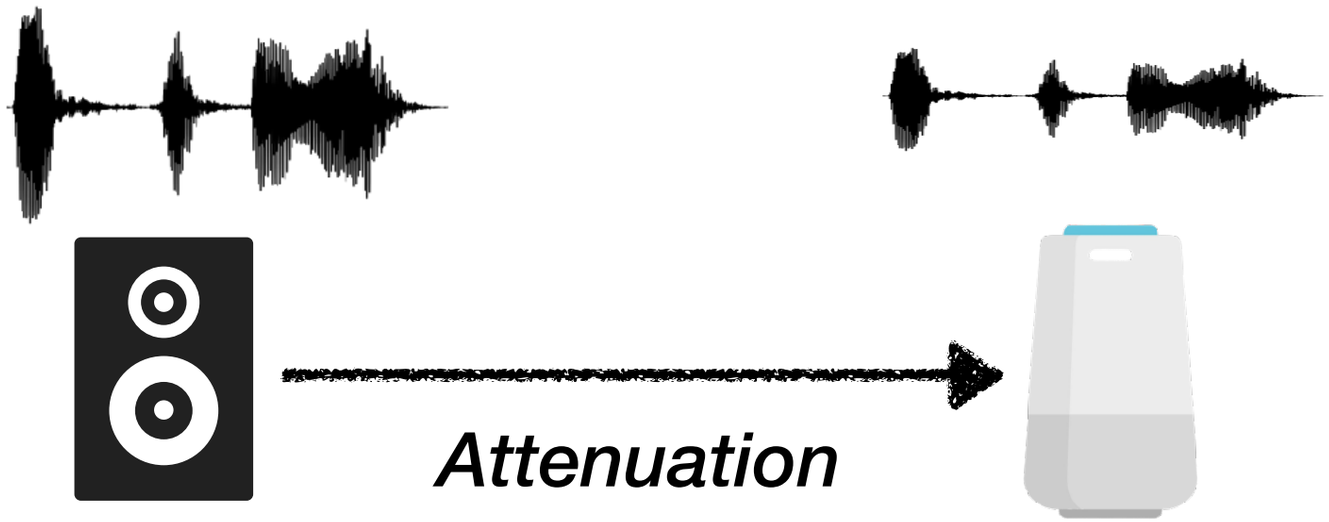
Understand Over-the-air Attack

Channel Effect

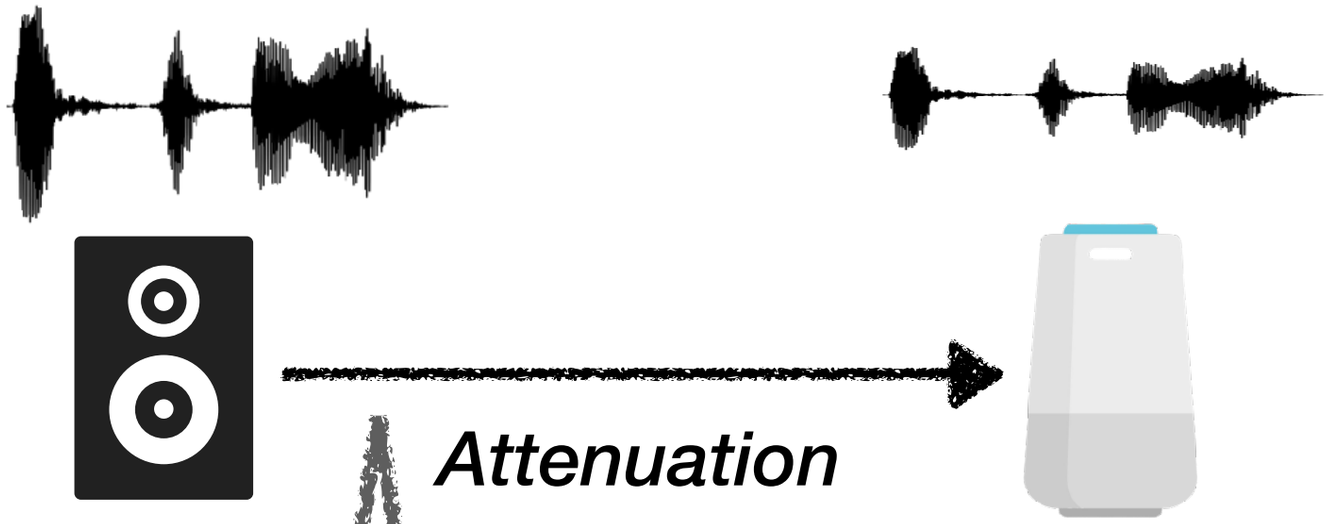


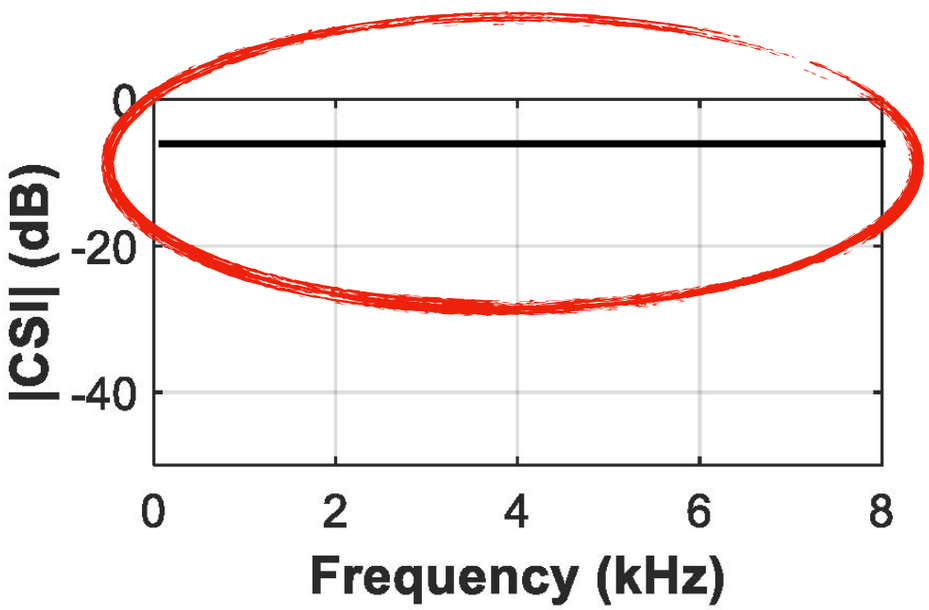
Hardware Heterogeneity

Attenuation



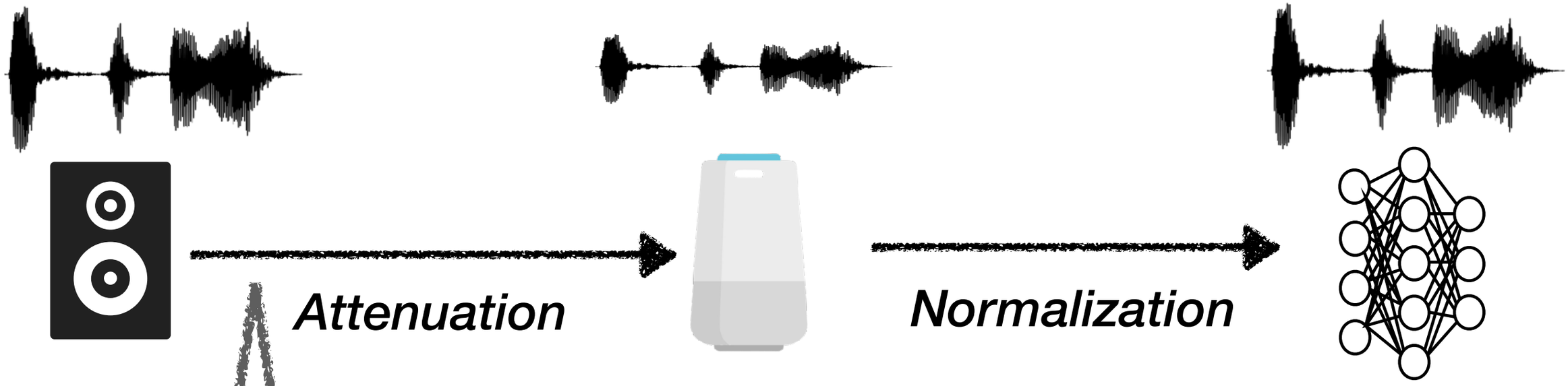
Attenuation

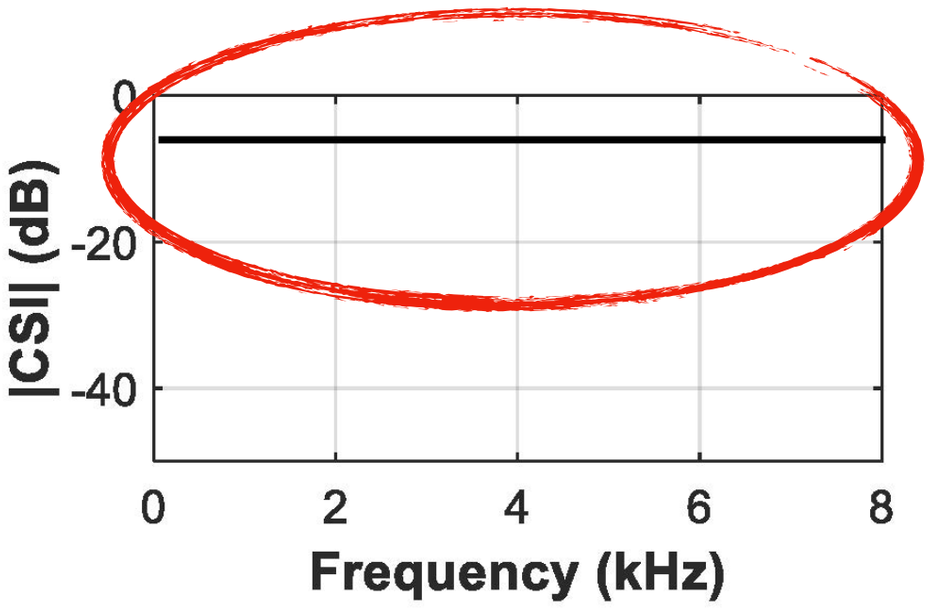




Attenuation

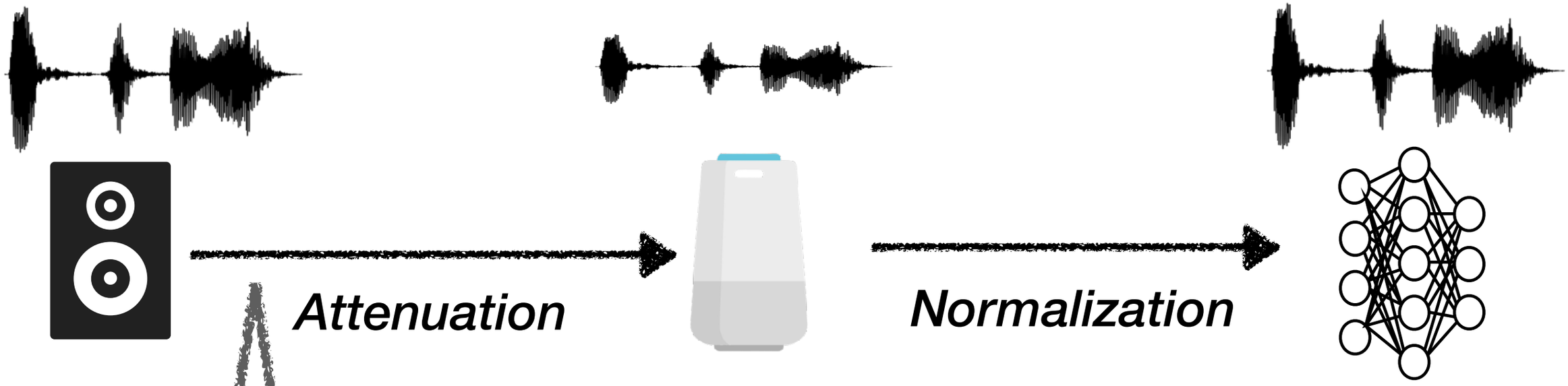
“ Open the door”

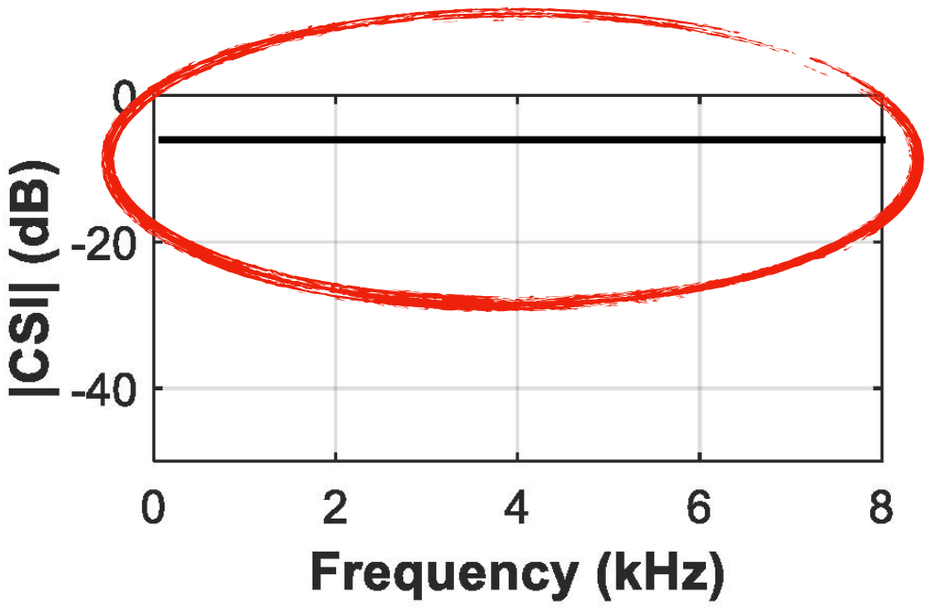




Attenuation

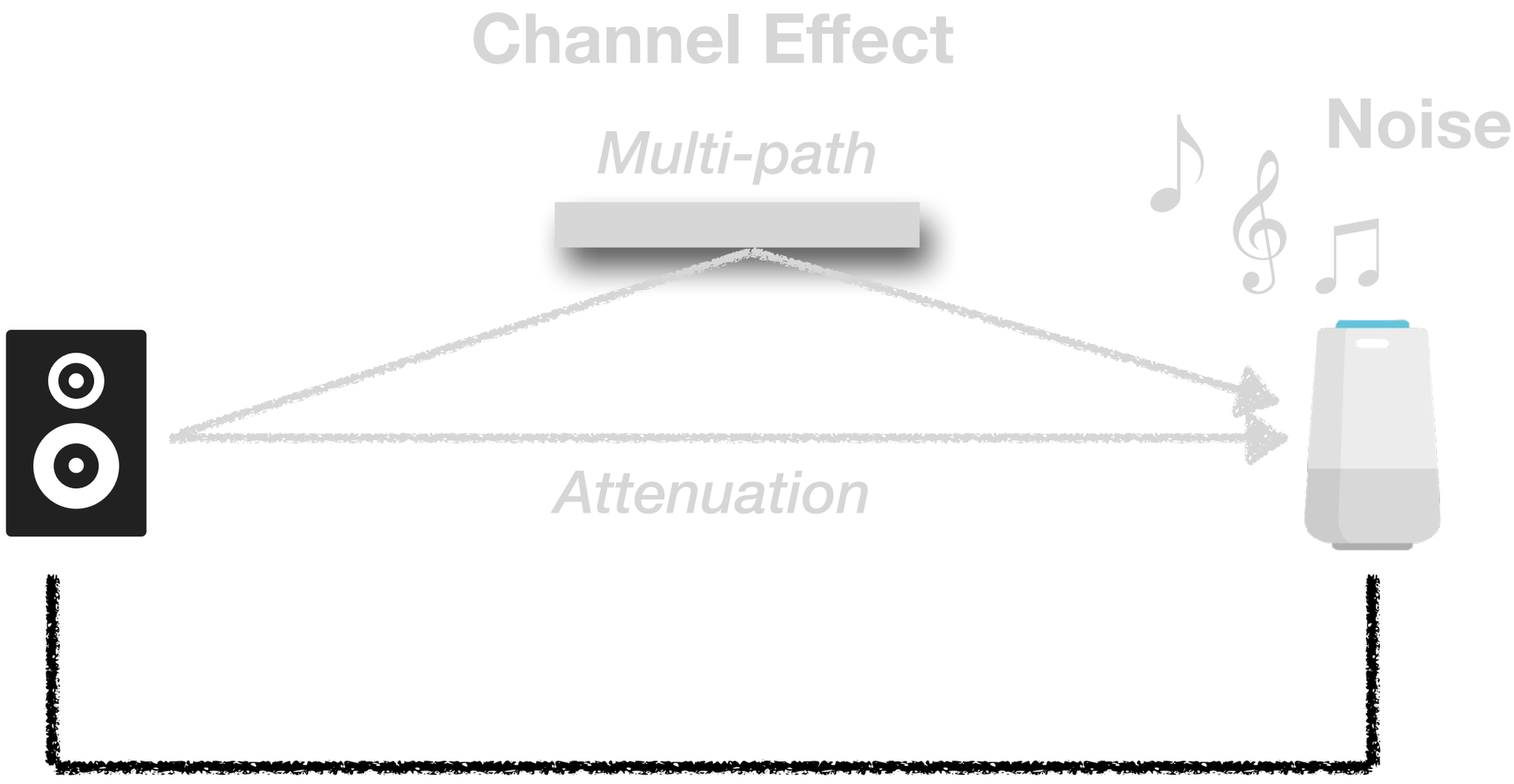
“ Open the door”





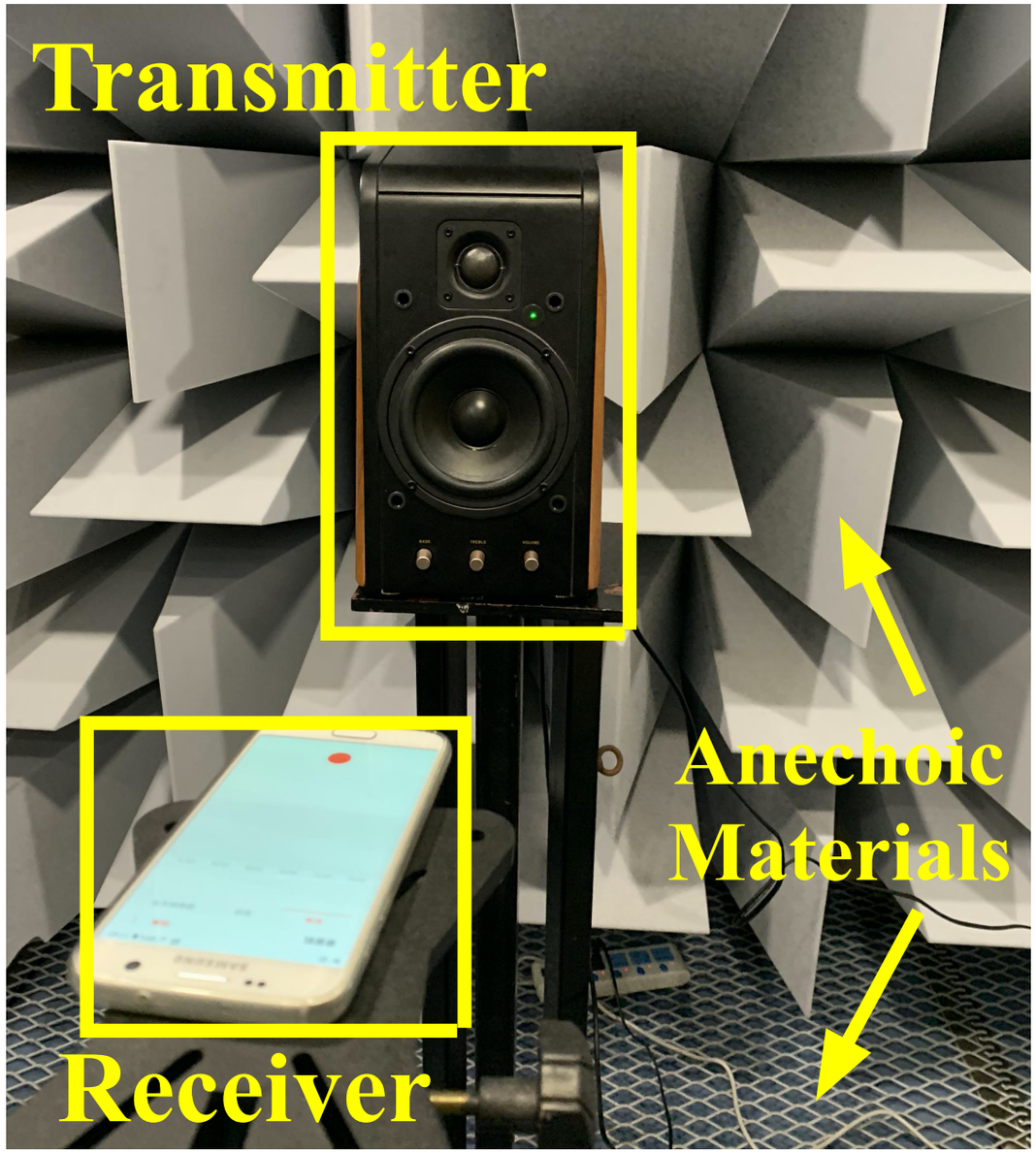
No frequency-selectivity, doesn’t matter at all!

Understand Over-the-air Attack



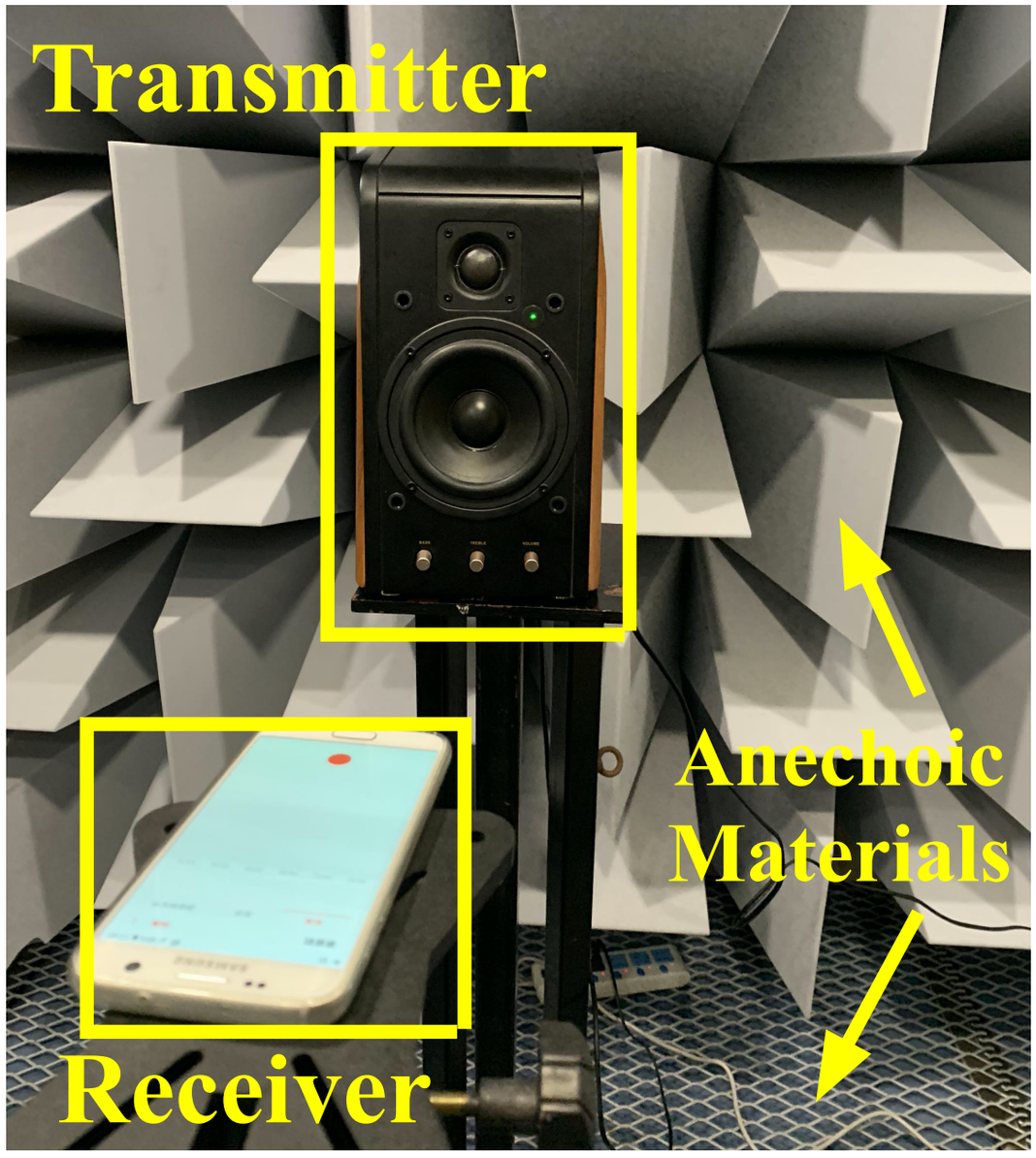
Hardware Heterogeneity

Hardware Heterogeneity

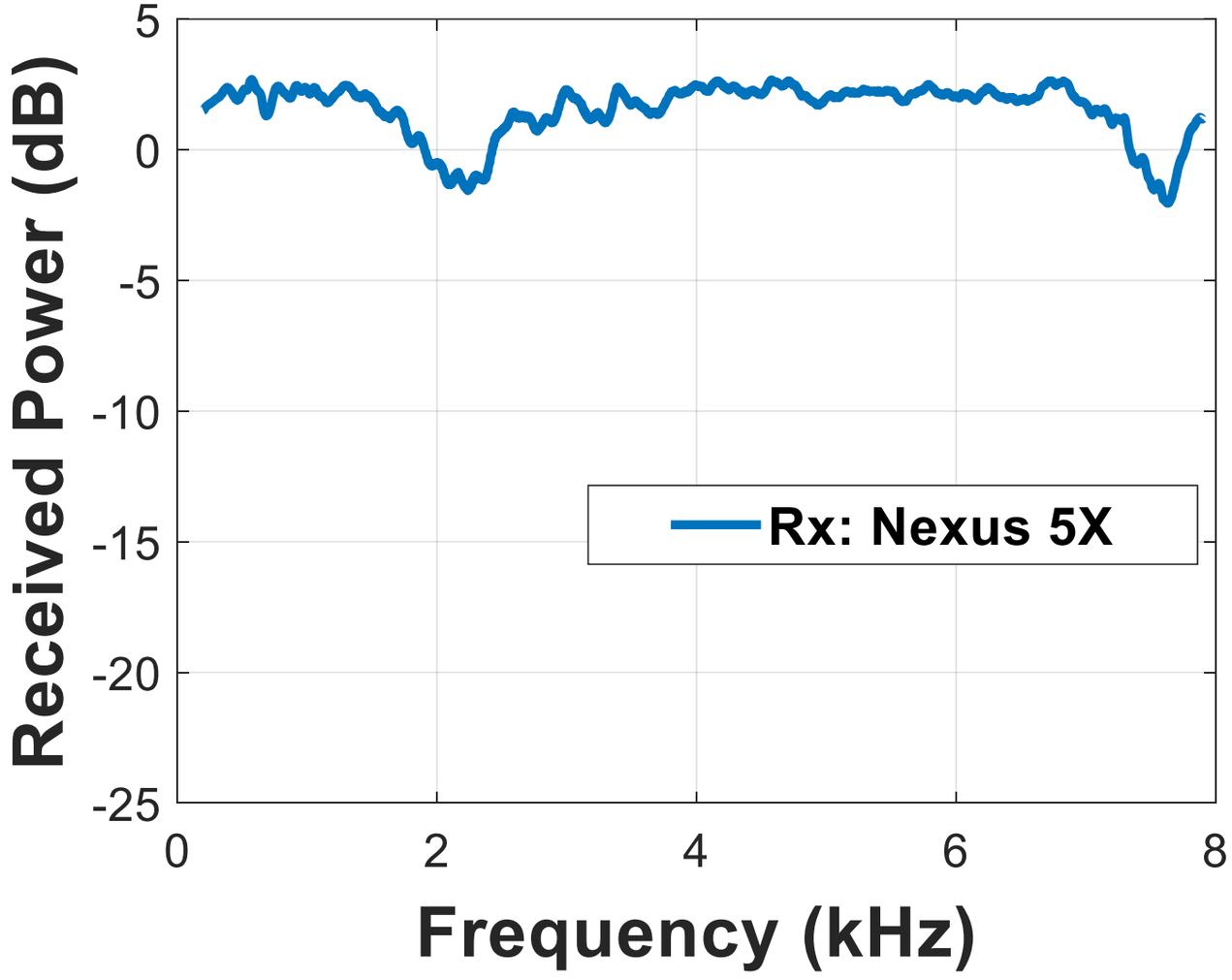


Anechoic Chamber Testing

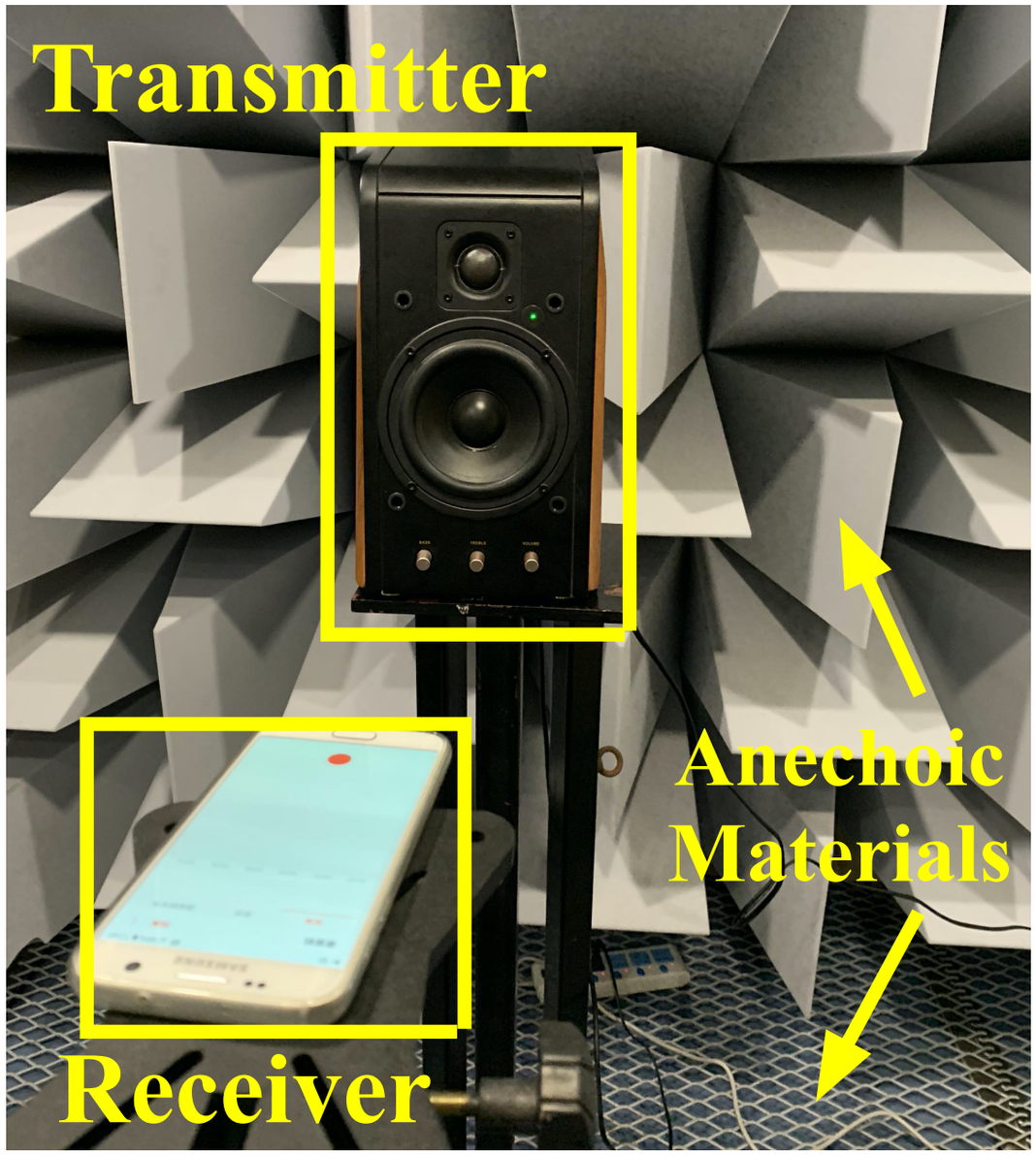
Hardware Heterogeneity



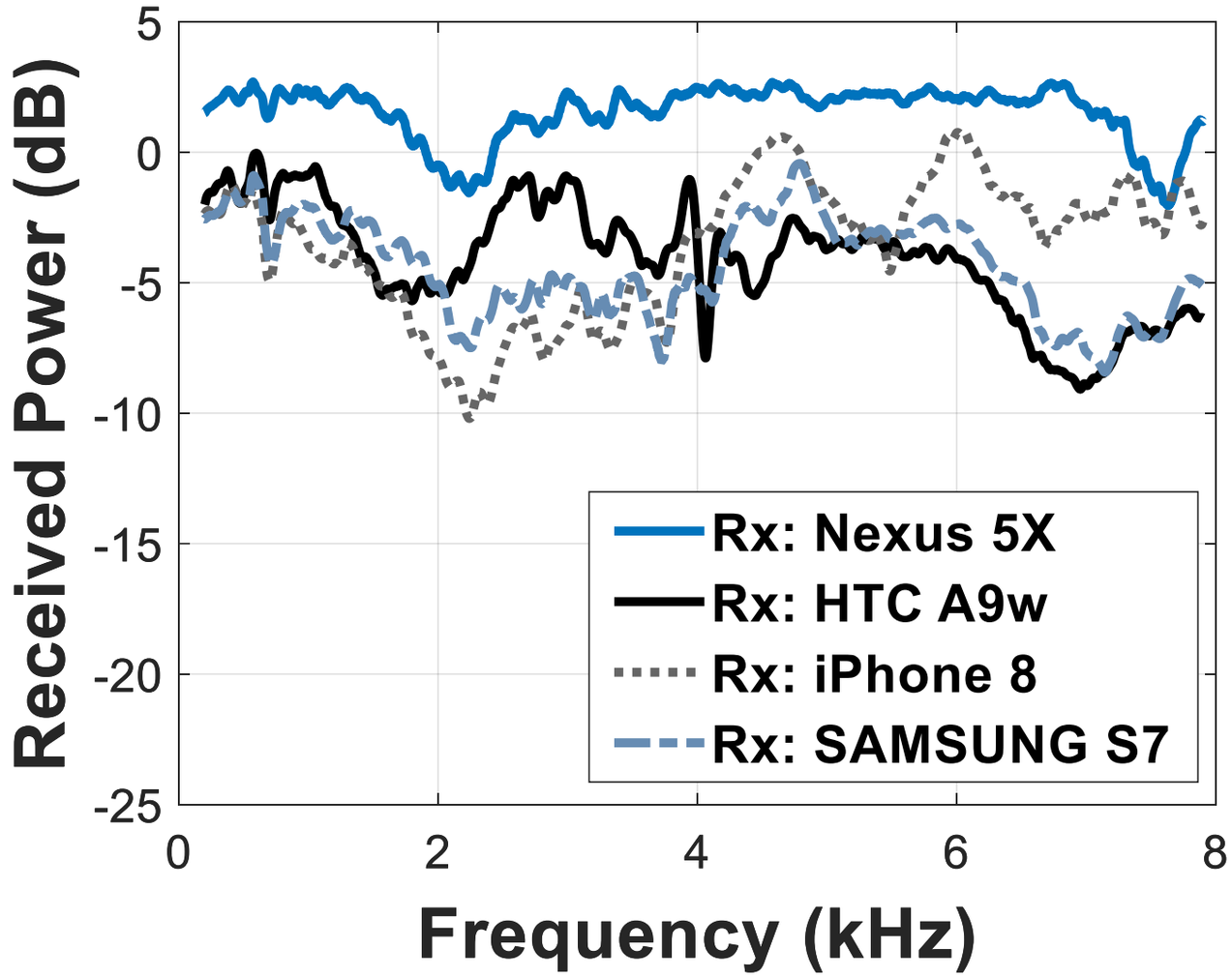
Anechoic Chamber Testing



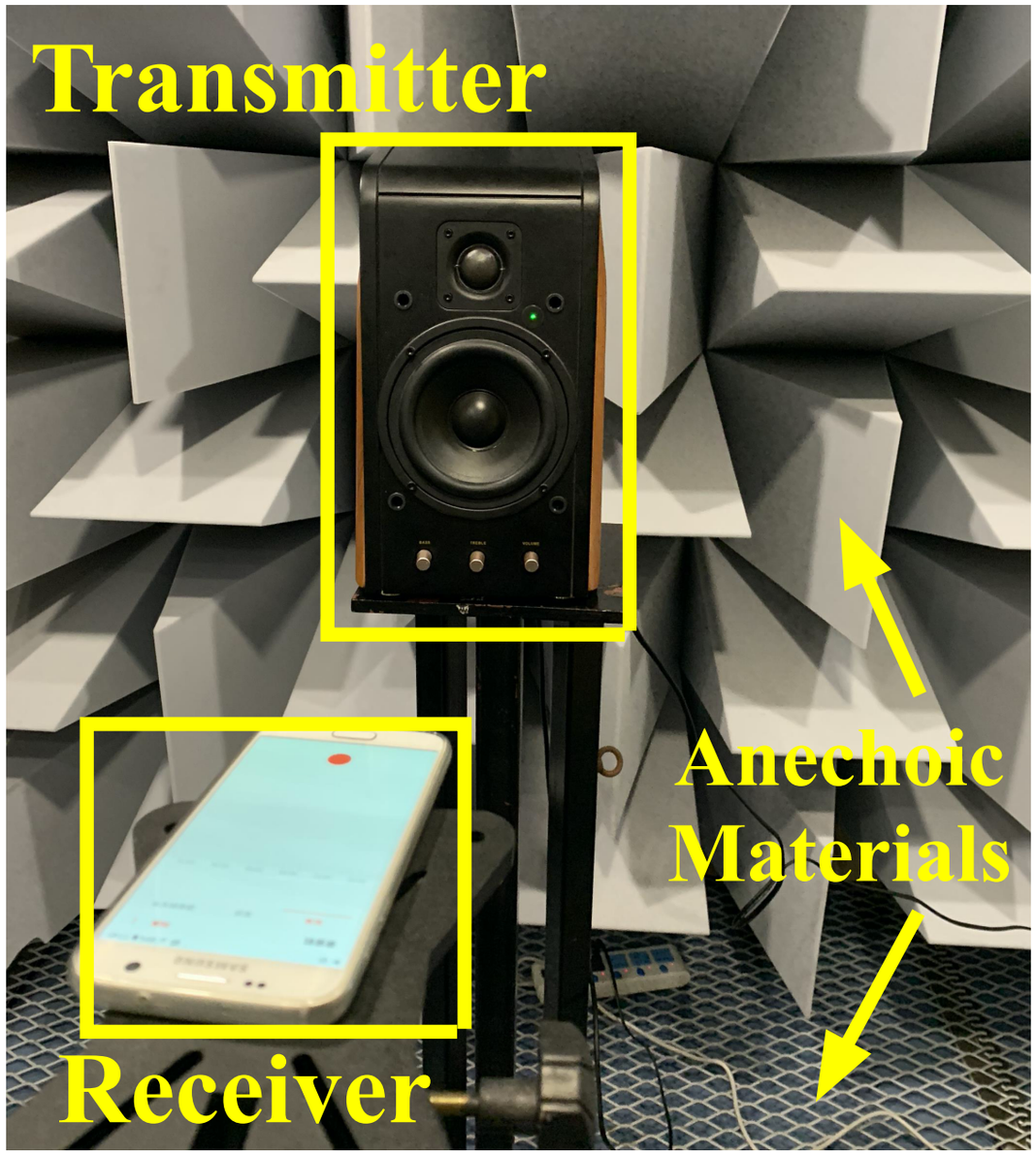
Hardware Heterogeneity



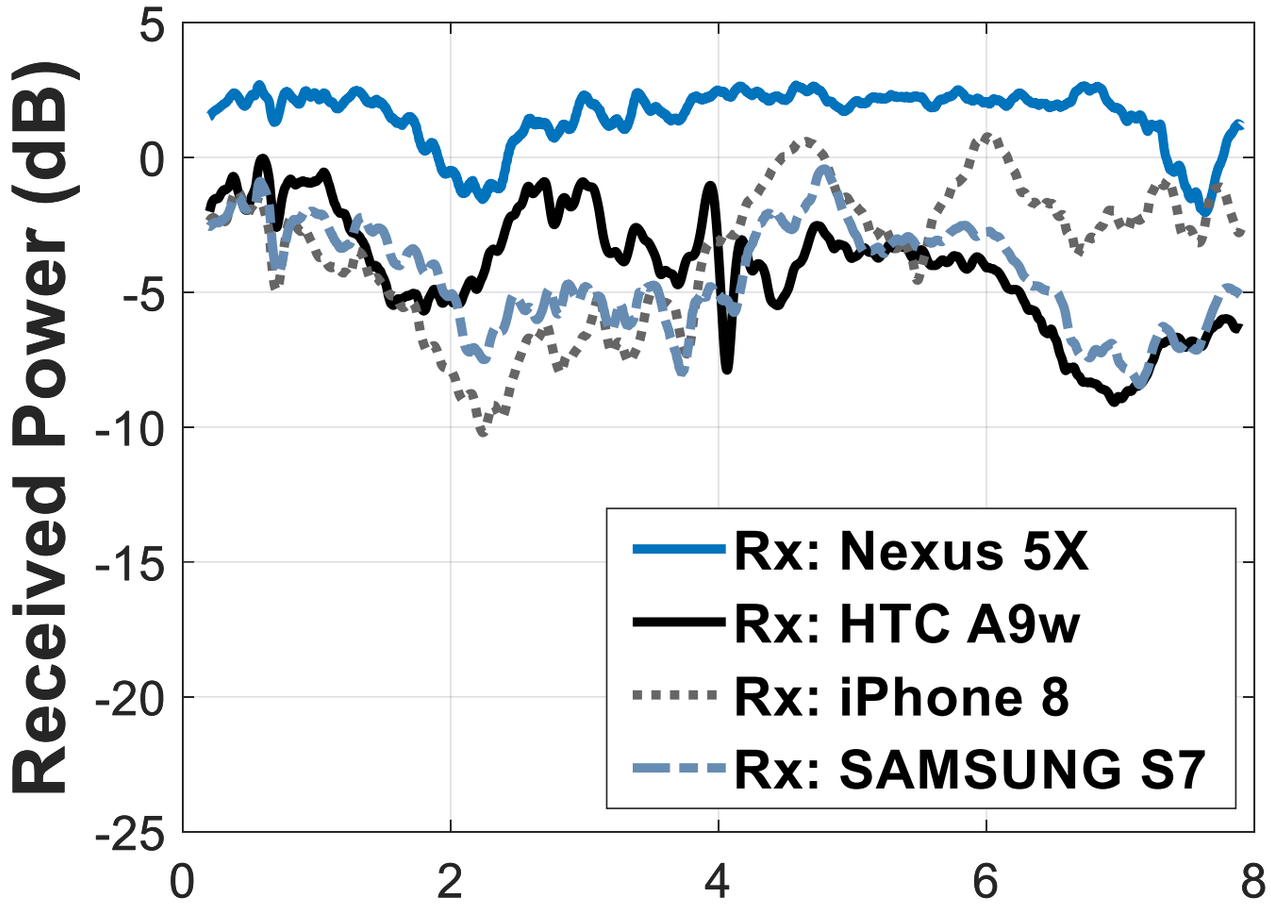
Anechoic Chamber Testing



Hardware Heterogeneity

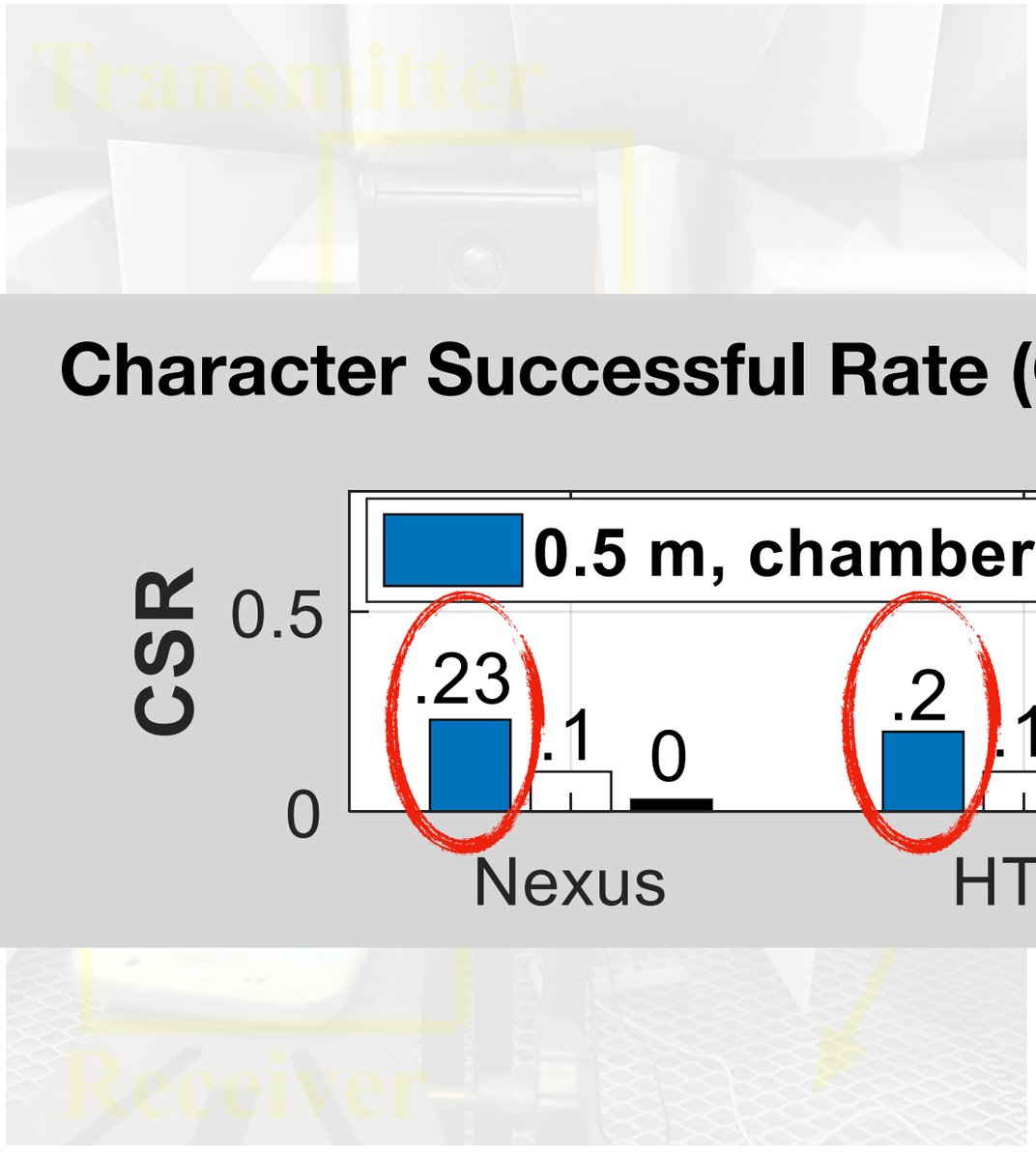


Anechoic Chamber Testing

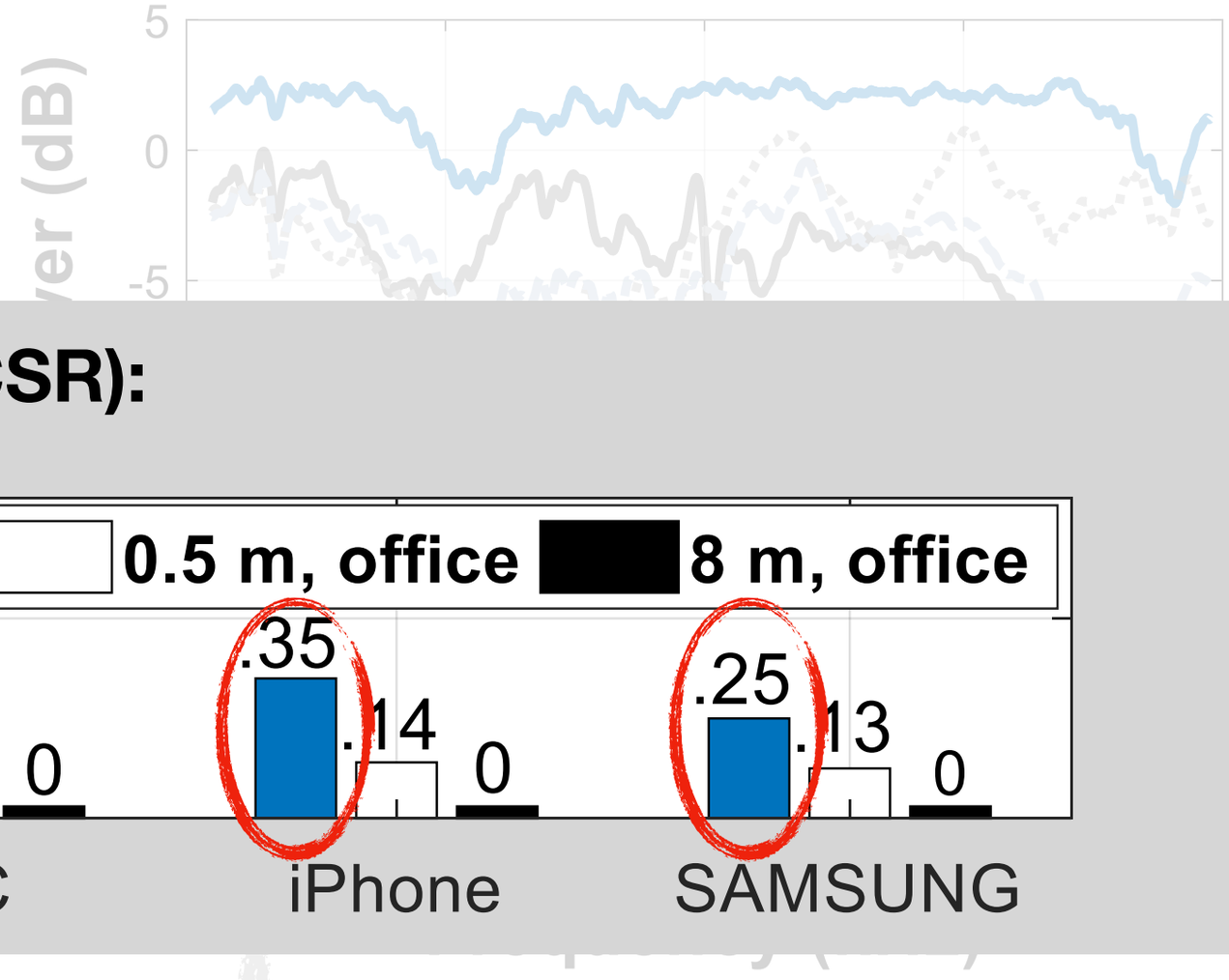


Not strong, device's inherent feature, compensable!

Hardware Heterogeneity



Anechoic Chamber Testing



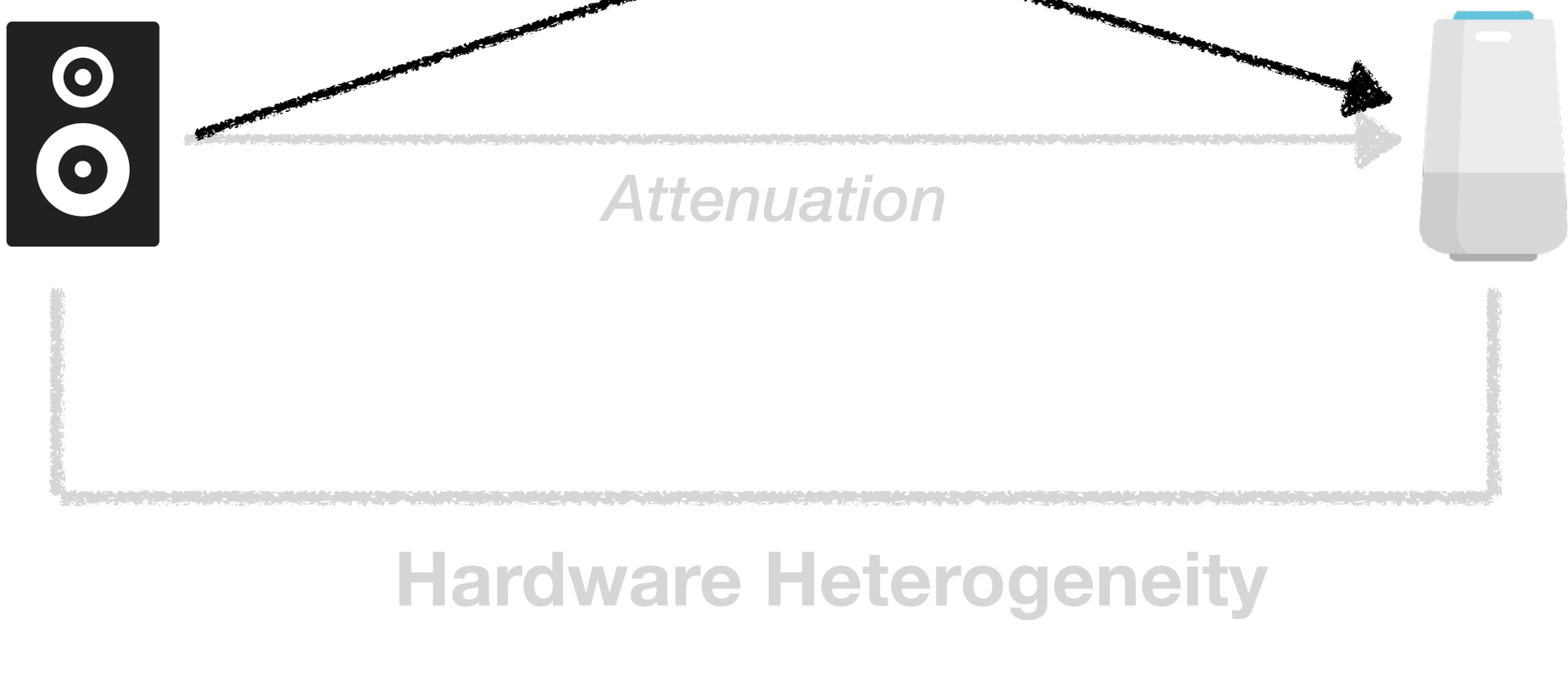
Character Successful Rate (CSR):

Static, predictable
and compensable!

Understand Over-the-air Attack

Channel Effect

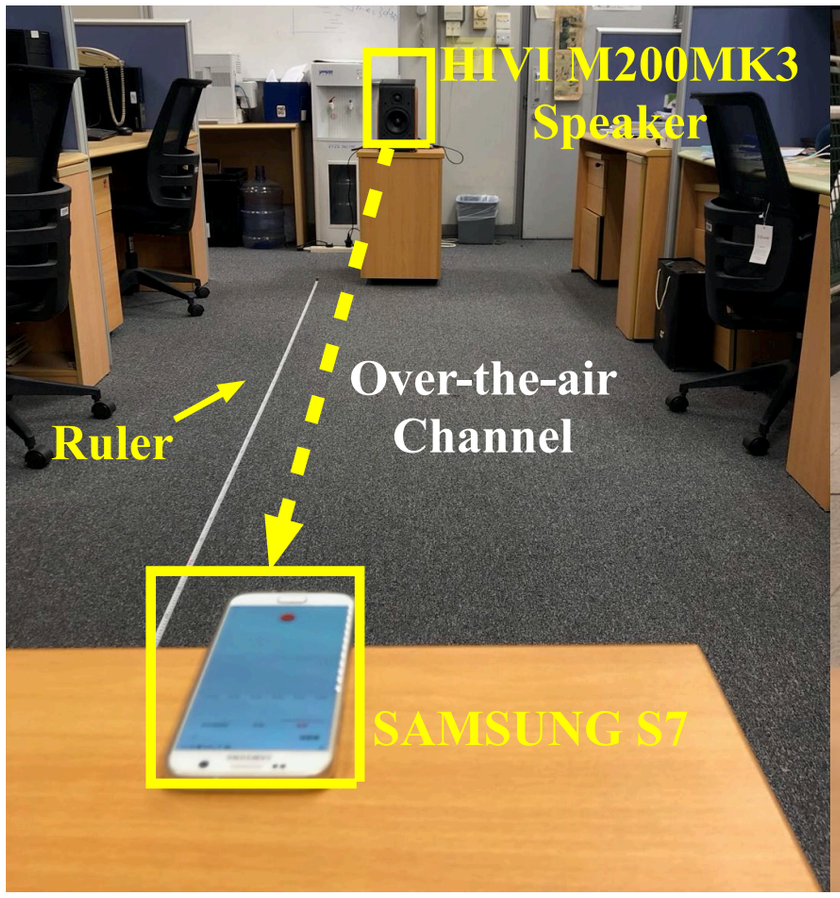
Multi-path

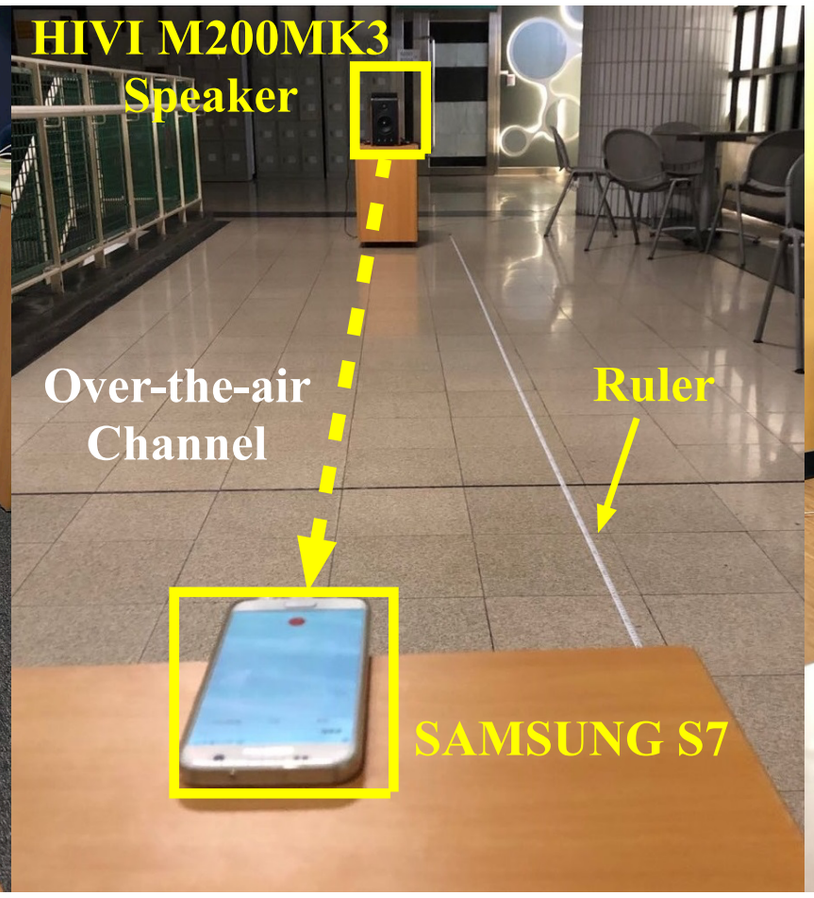


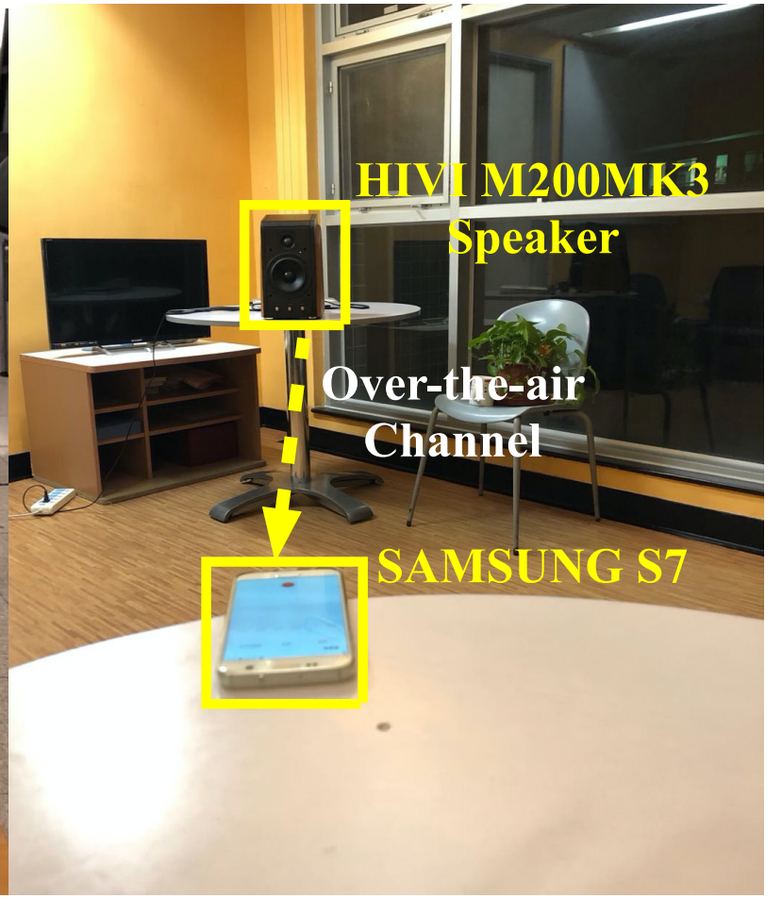
Attenuation

Hardware Heterogeneity

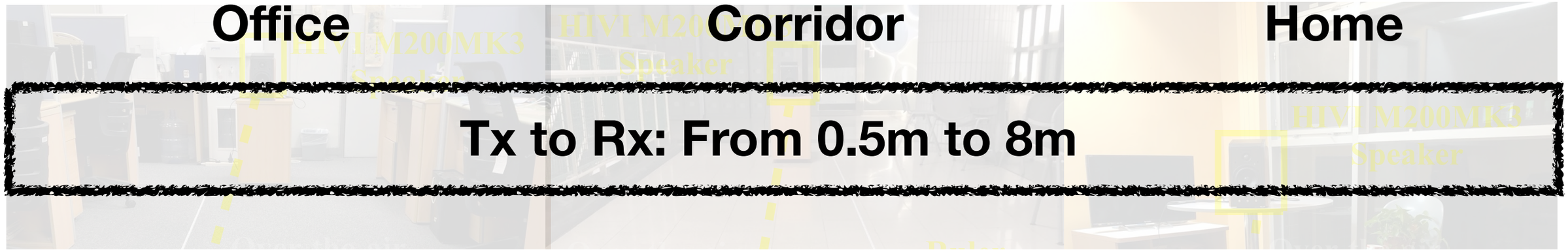
Multi-path


HIVI M200MK3
Speaker
Over-the-air
Channel
Ruler
SAMSUNG S7


HIVI M200MK3
Speaker
Over-the-air
Channel
Ruler
SAMSUNG S7


HIVI M200MK3
Speaker
Over-the-air
Channel
SAMSUNG S7

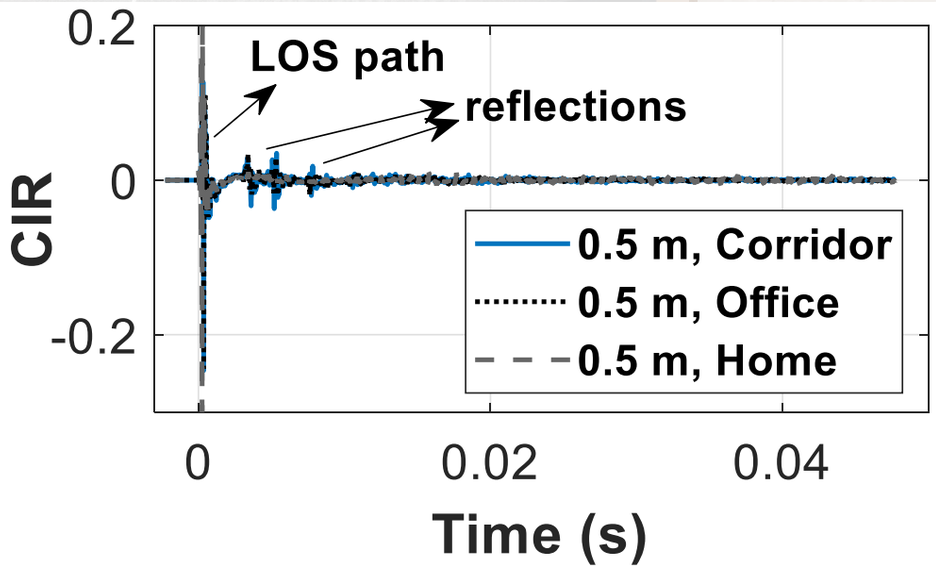
Multi-path: Near range

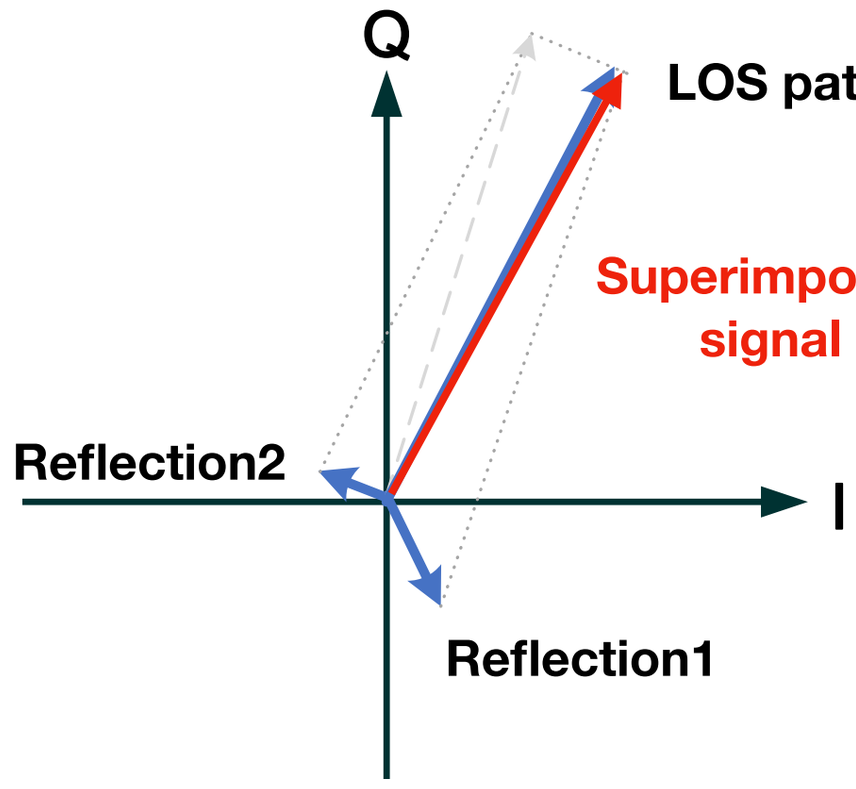


Multi-path: Near range

Office **Corridor** **Home**

Tx to Rx: From 0.5m to 8m

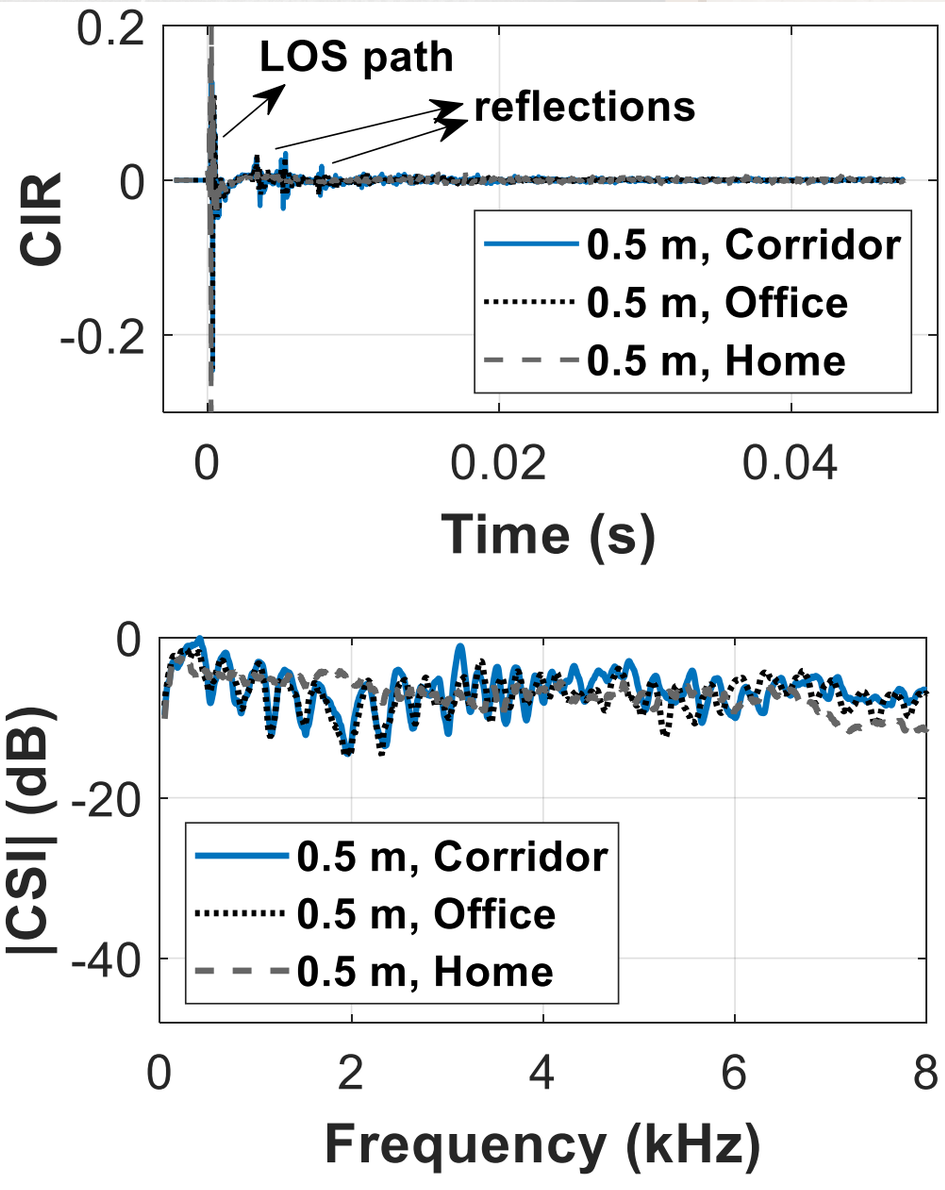


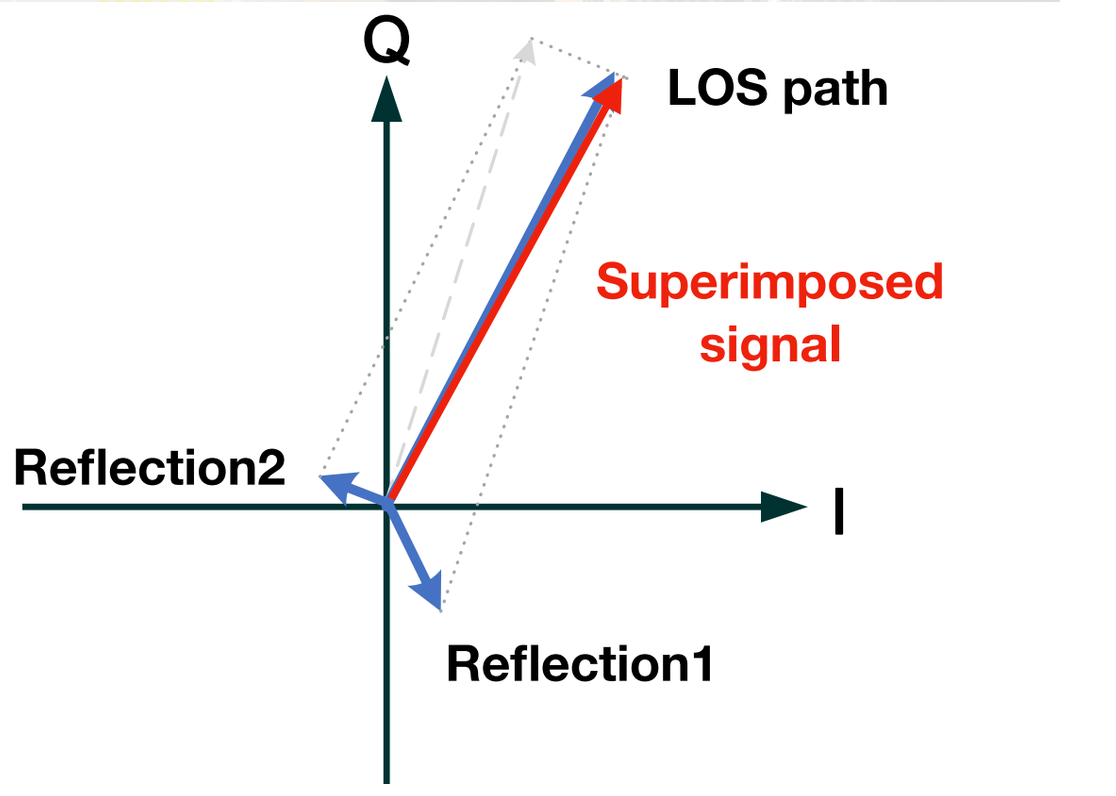


Multi-path: Near range

Office **Corridor** **Home**

Tx to Rx: From 0.5m to 8m

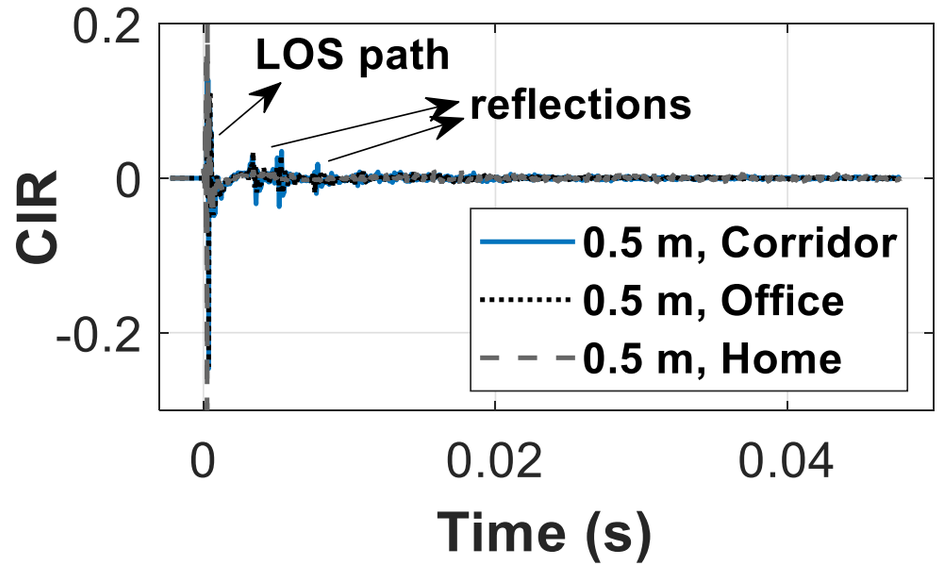


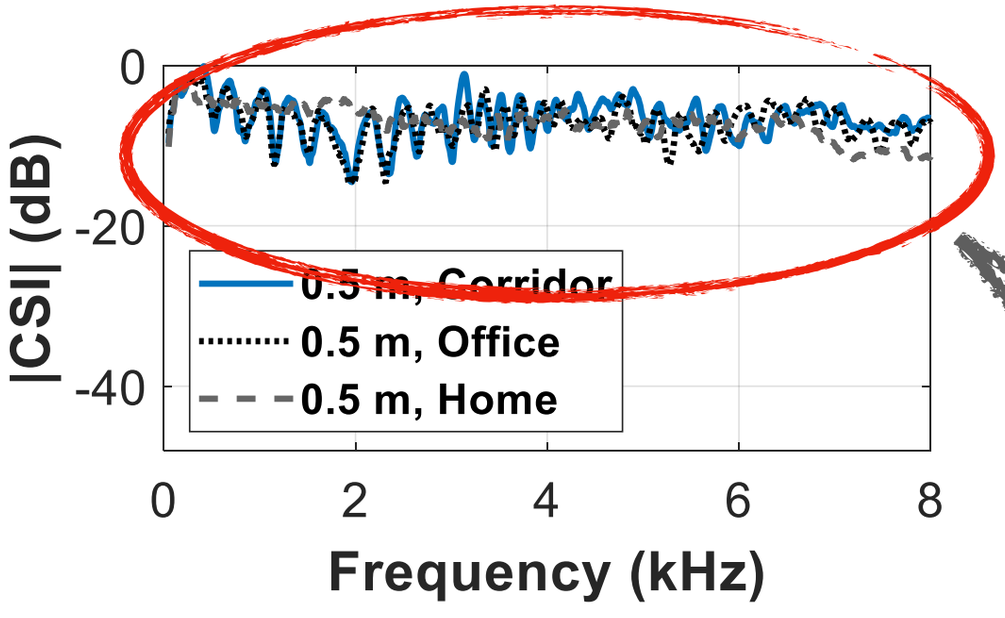


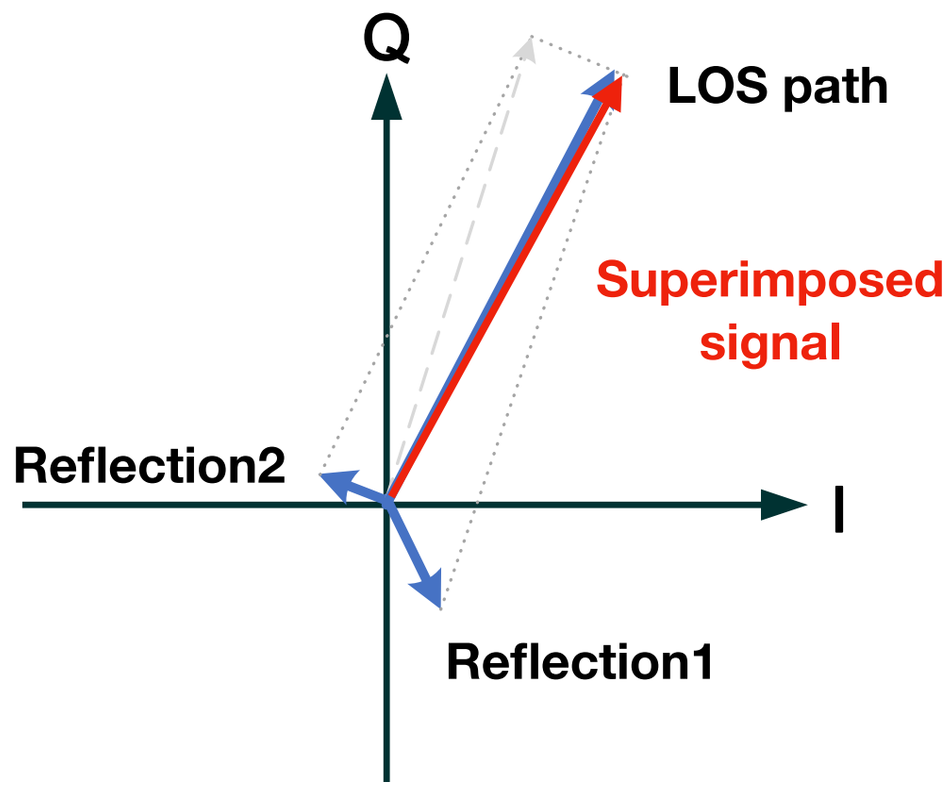
Multi-path: Near range

Office **Corridor** **Home**

Tx to Rx: From 0.5m to 8m

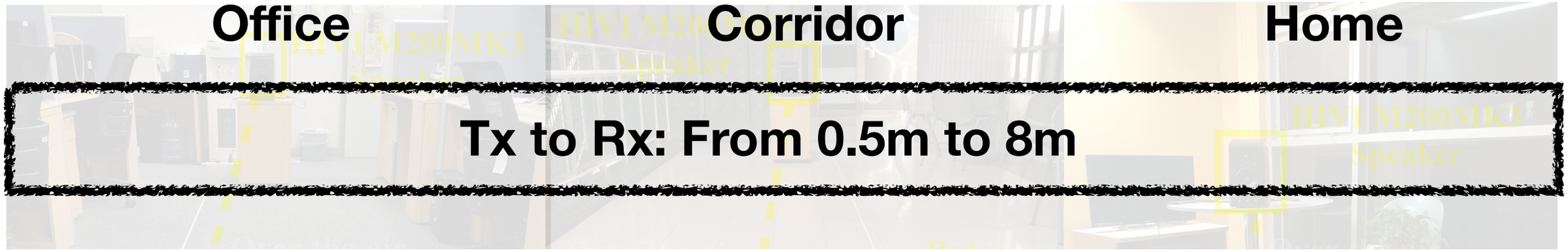






Also not strong and similar!

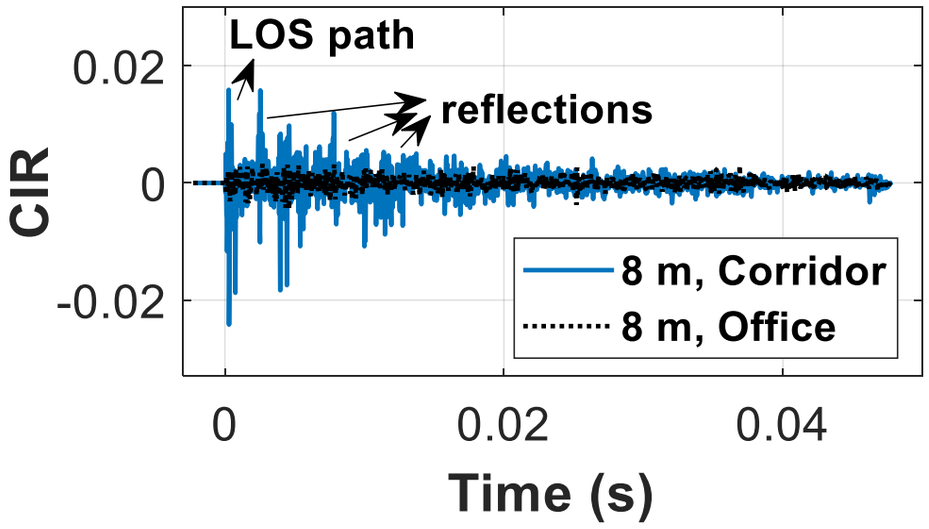
Multi-path: Long range

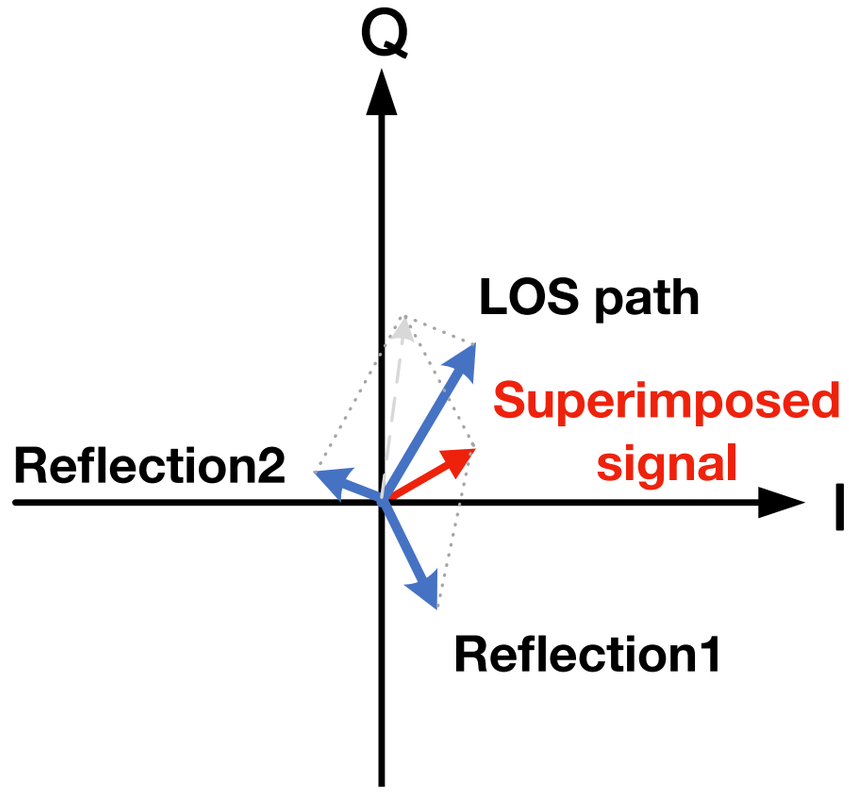


Multi-path: Long range

Office **Corridor** **Home**

Tx to Rx: From 0.5m to 8m

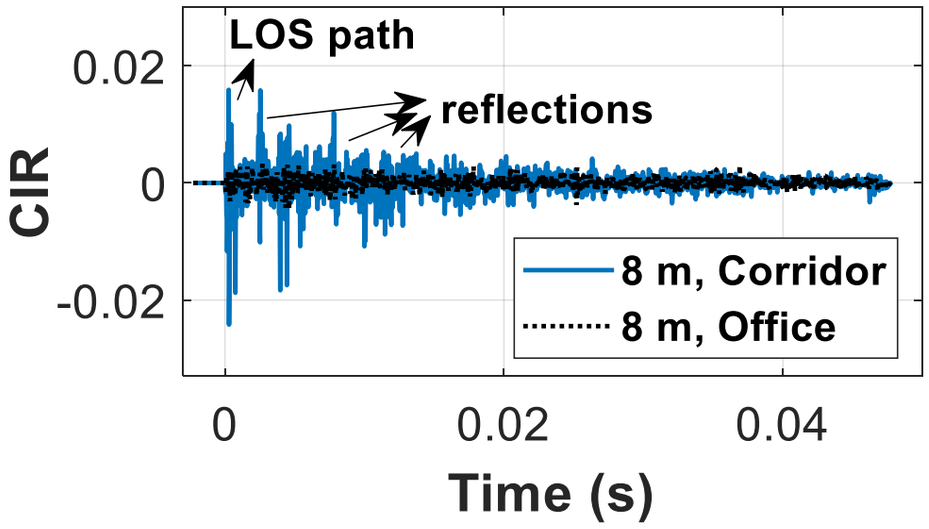


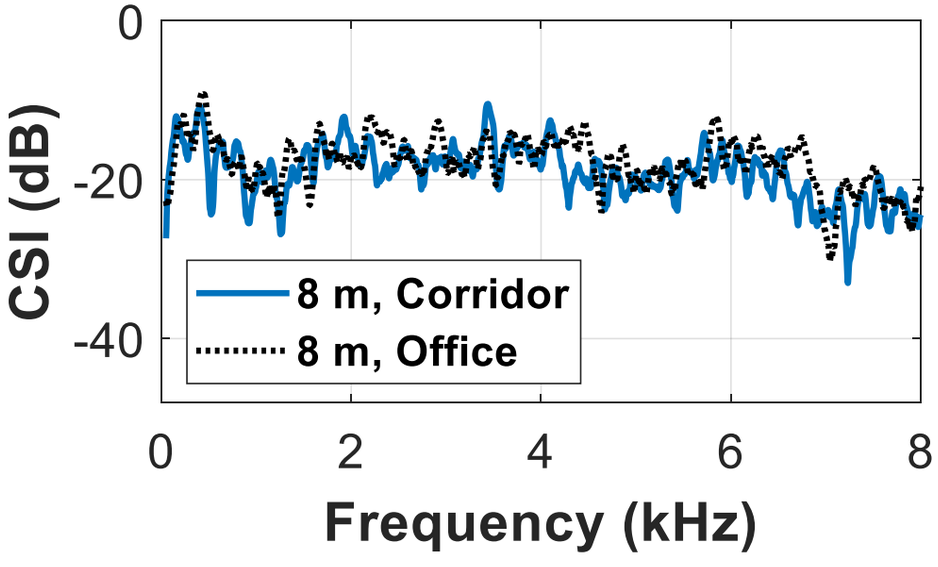


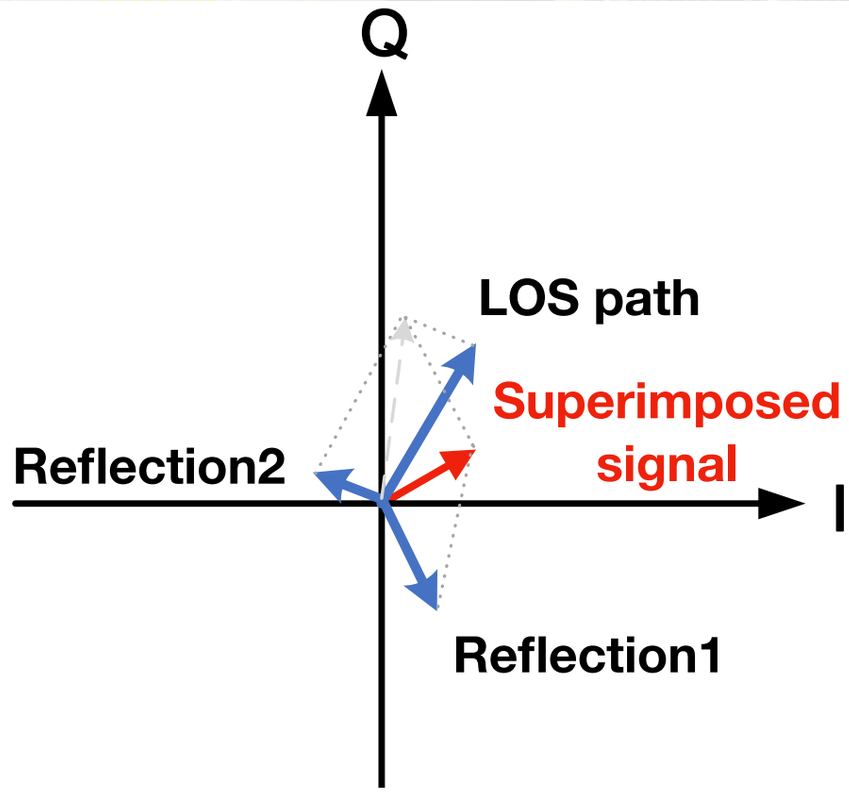
Multi-path: Long range

Office **Corridor** **Home**

Tx to Rx: From 0.5m to 8m



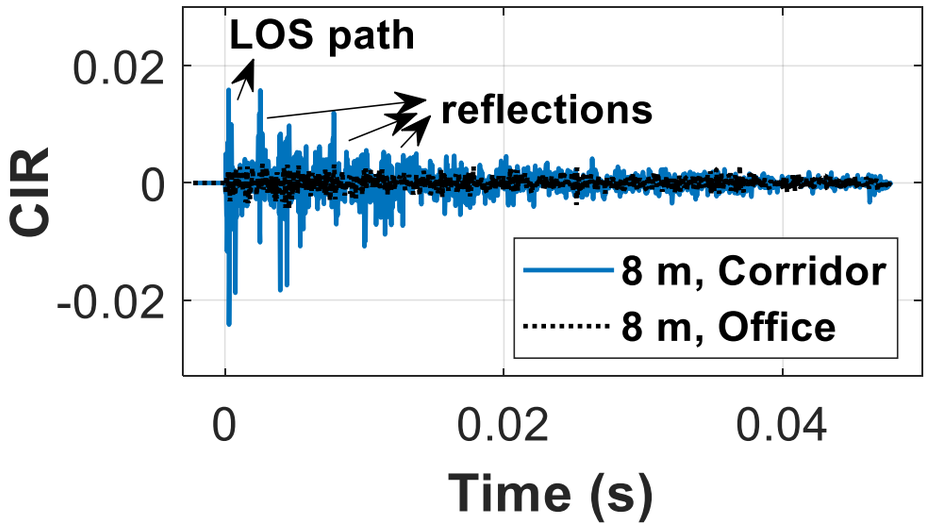


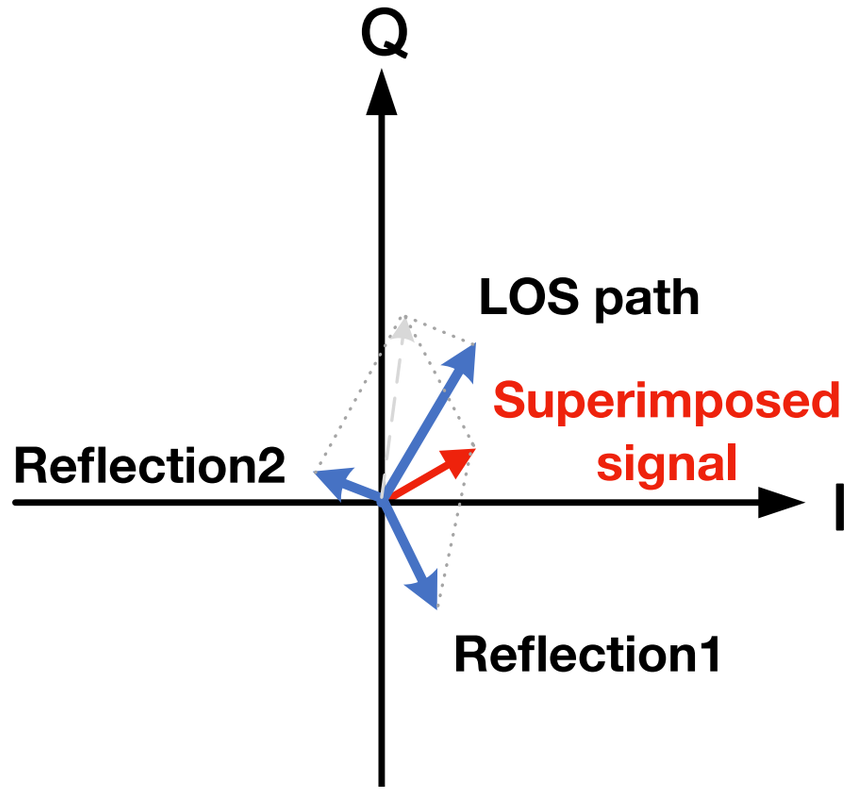


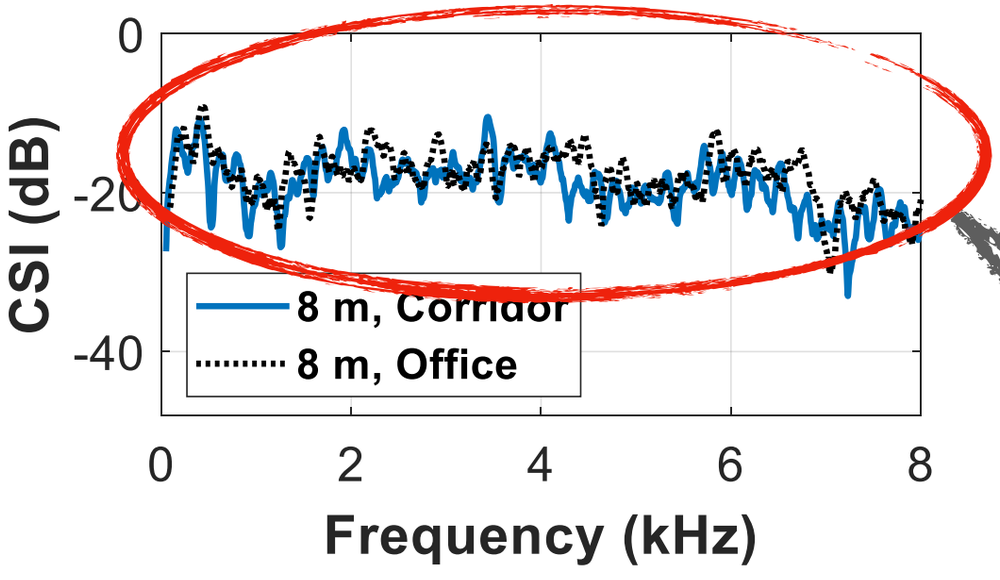
Multi-path: Long range

Office **Corridor** **Home**

Tx to Rx: From 0.5m to 8m



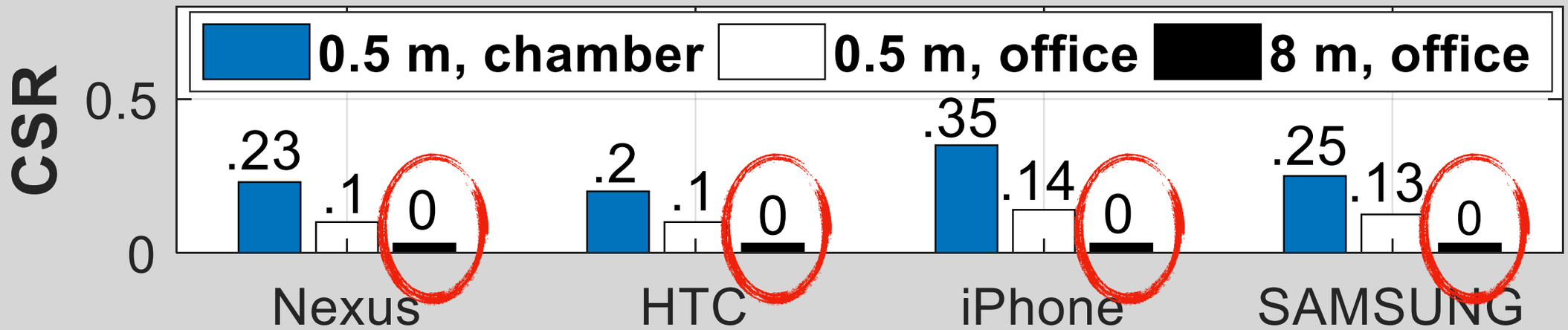


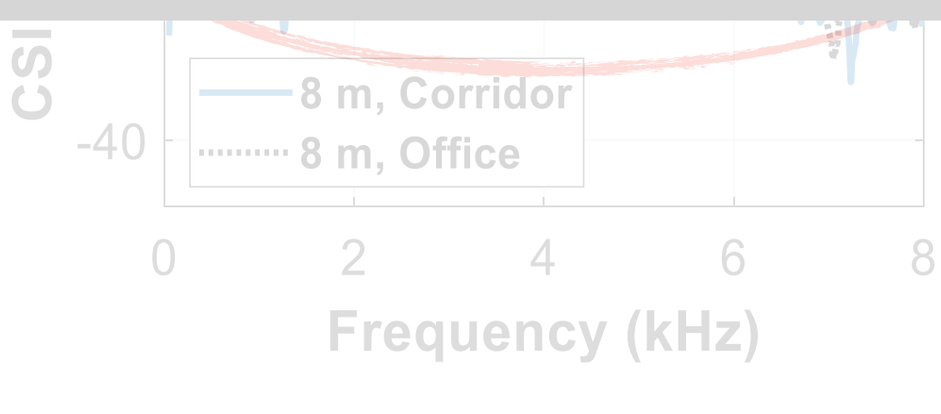


Stronger and unpredictable!

Multi-path: Long range

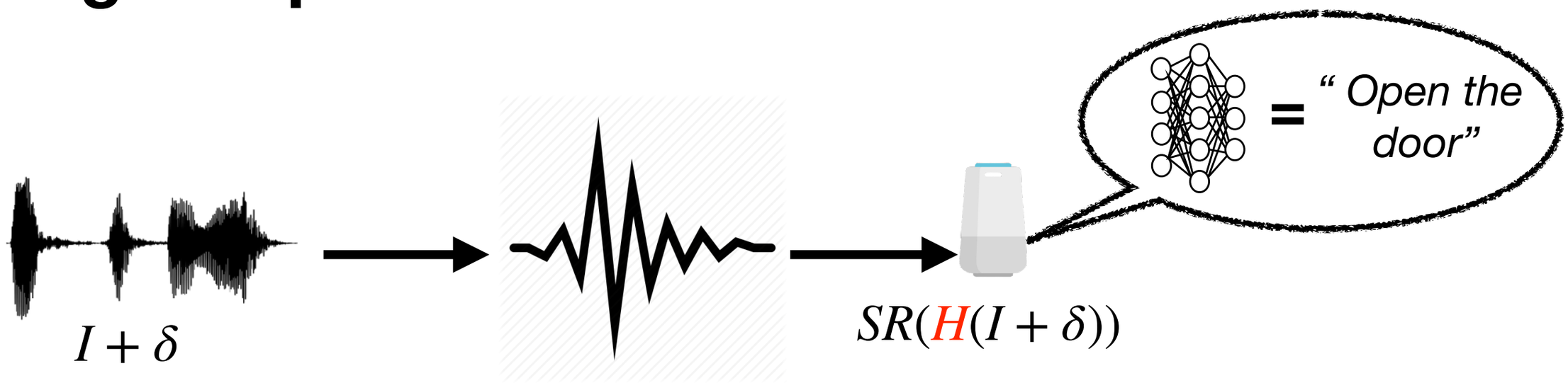
Character Successful Rate (CSR):





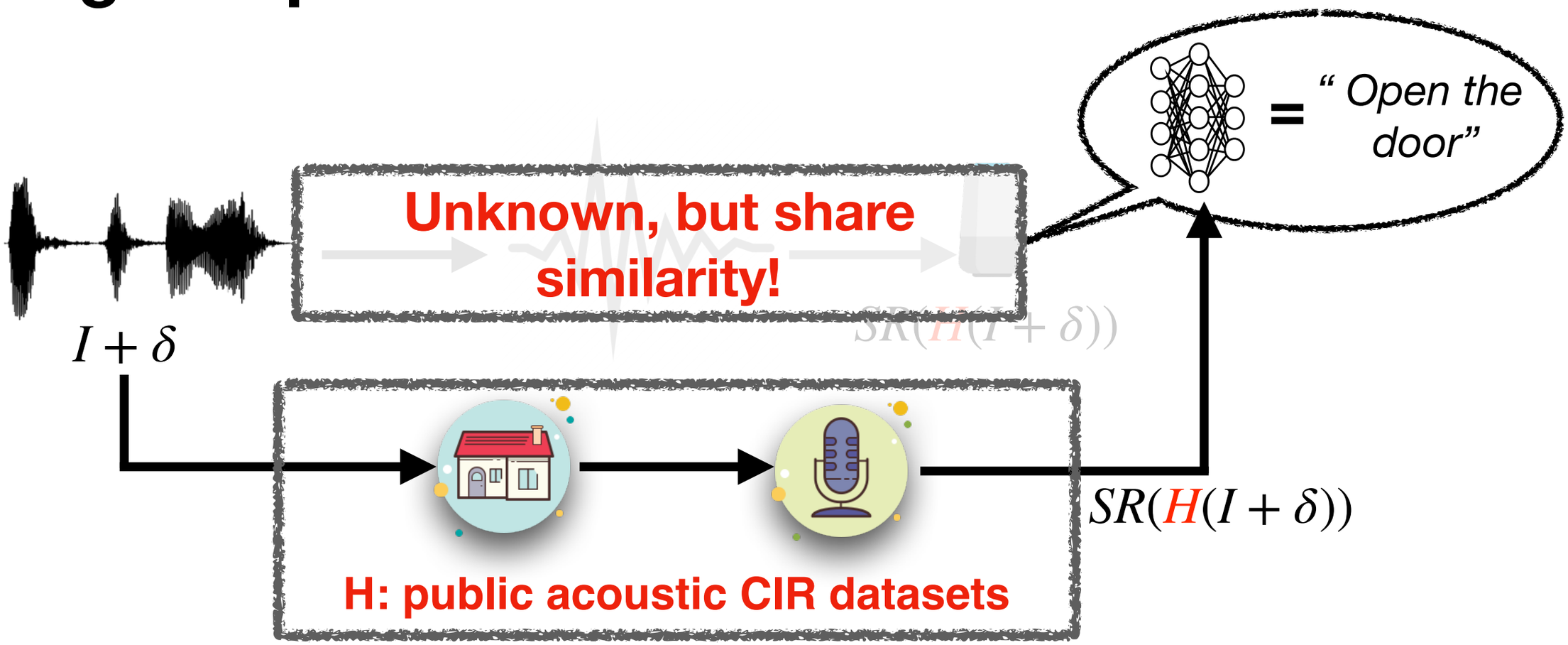
Highly unpredictable!

Design Inspiration

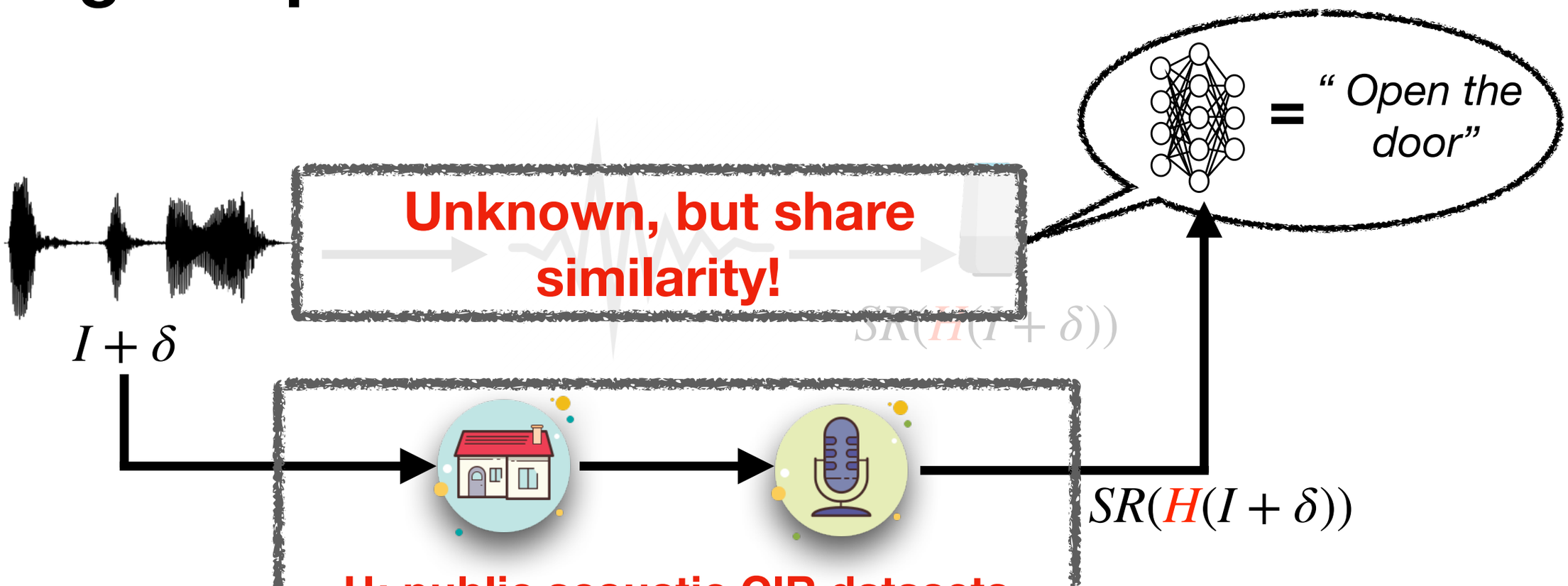


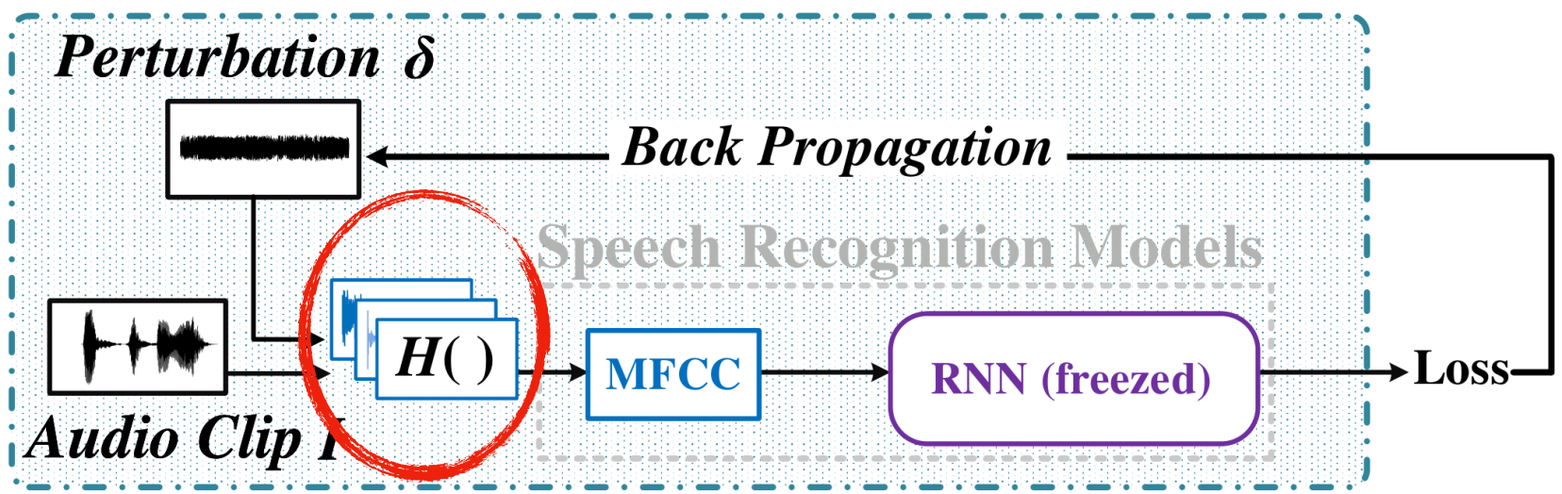
Design Inspiration

Design Inspiration



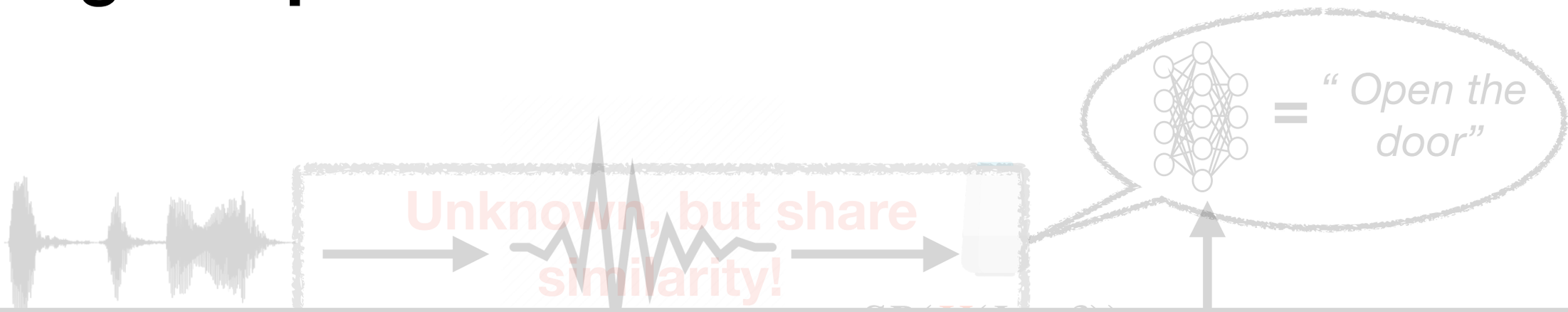
Design Inspiration



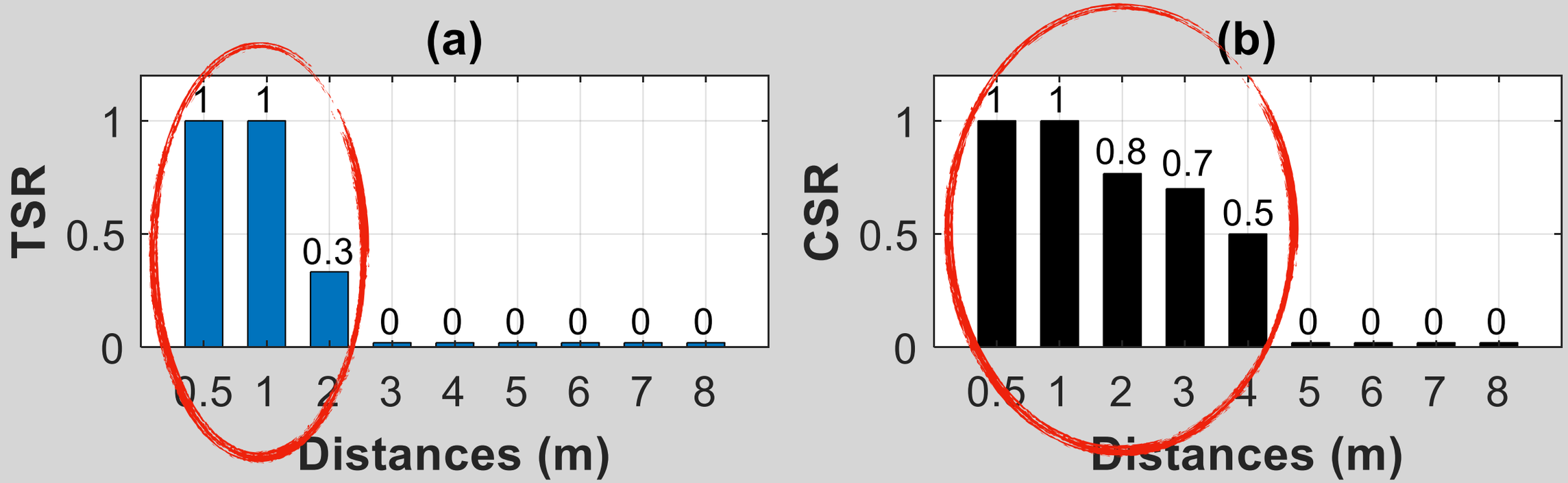


$$\arg\min_{\delta} \ \alpha \cdot dB_I(\delta) + \frac{1}{M}\sum_i Loss(SR(H_i(I+\delta)), T')$$

Design Inspiration

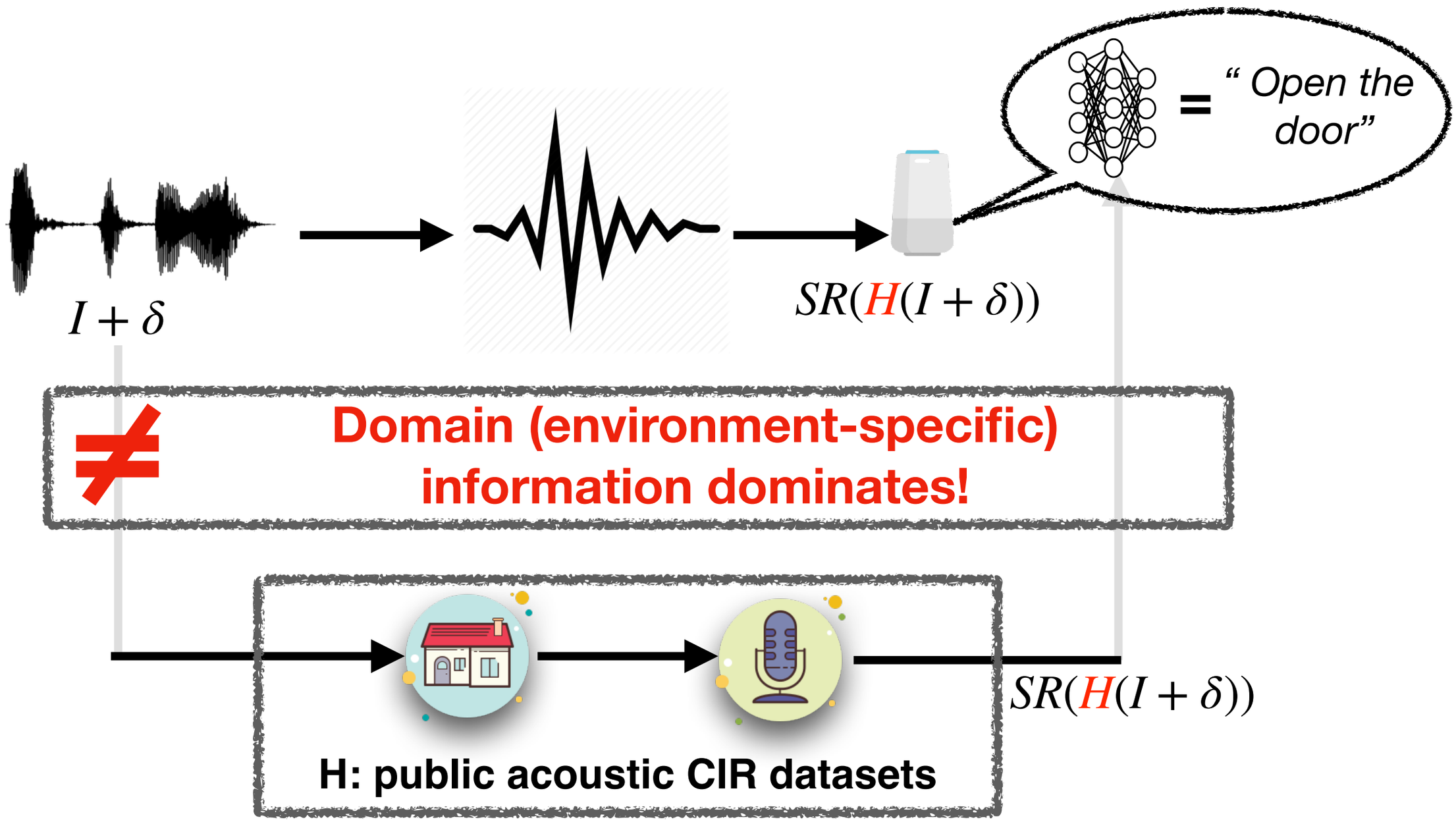


Transcript and Character Successful Rate:

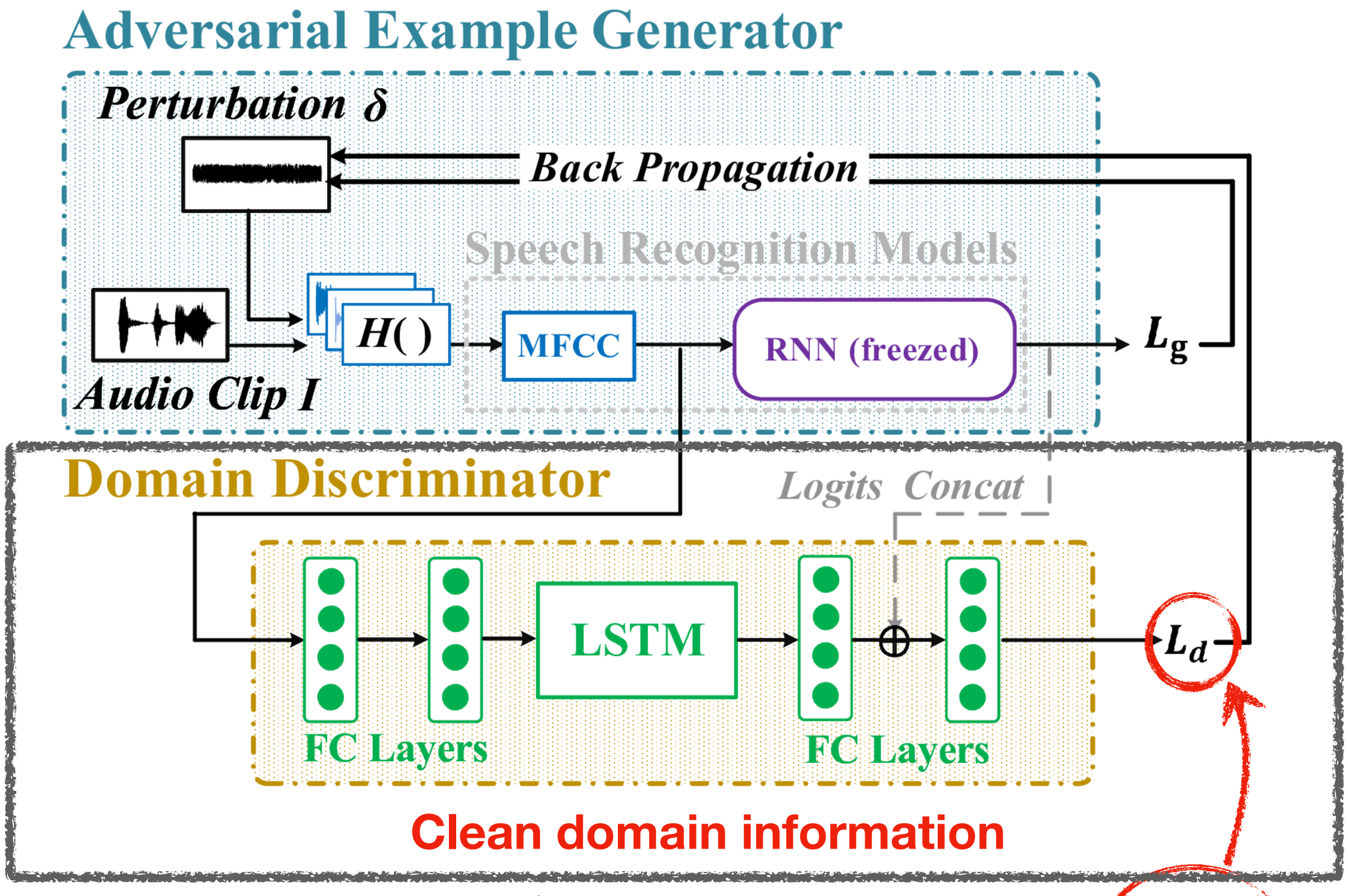


$$\arg\min_{\delta} \ \alpha \cdot dB_I(\delta) + \frac{1}{M}\sum_i Loss(SR(H_i(I+\delta)), T')$$

Design Inspiration



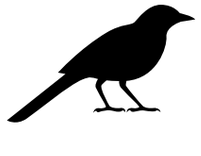
Metamorph: Meta-Enha



$$\arg\min_{\delta} \ \alpha \cdot dB_I(\delta) + \frac{1}{M}\sum_i Loss(SR(H_i(I+\delta)), T') - \beta \cdot L_d$$

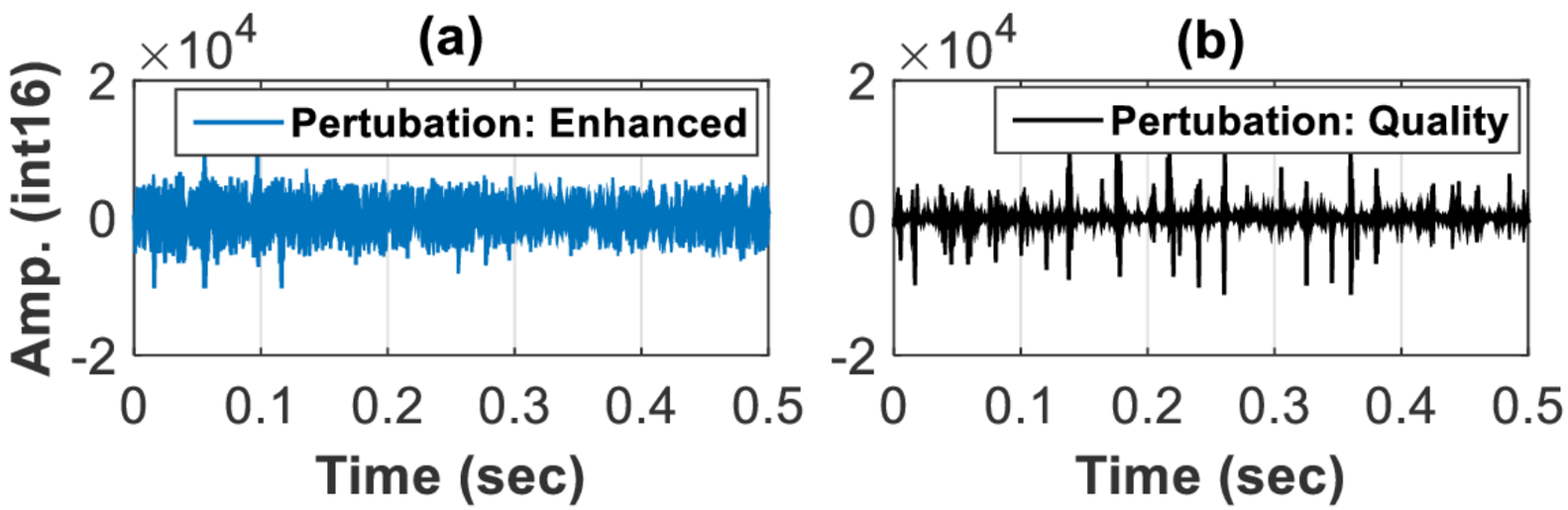
Metamorph: Meta-Qual

- Acoustic Graffiti:

$$distance(\delta, \hat{N})$$  

- Reducing Perturbation's Coverage:

L1/L2 regularization



Evaluation: Audio Quality

- Examples

Classical music

Original:
[no transcription]

Meta-Enha:
“hello world”

Meta-Qual:
“hello world”

Human speech

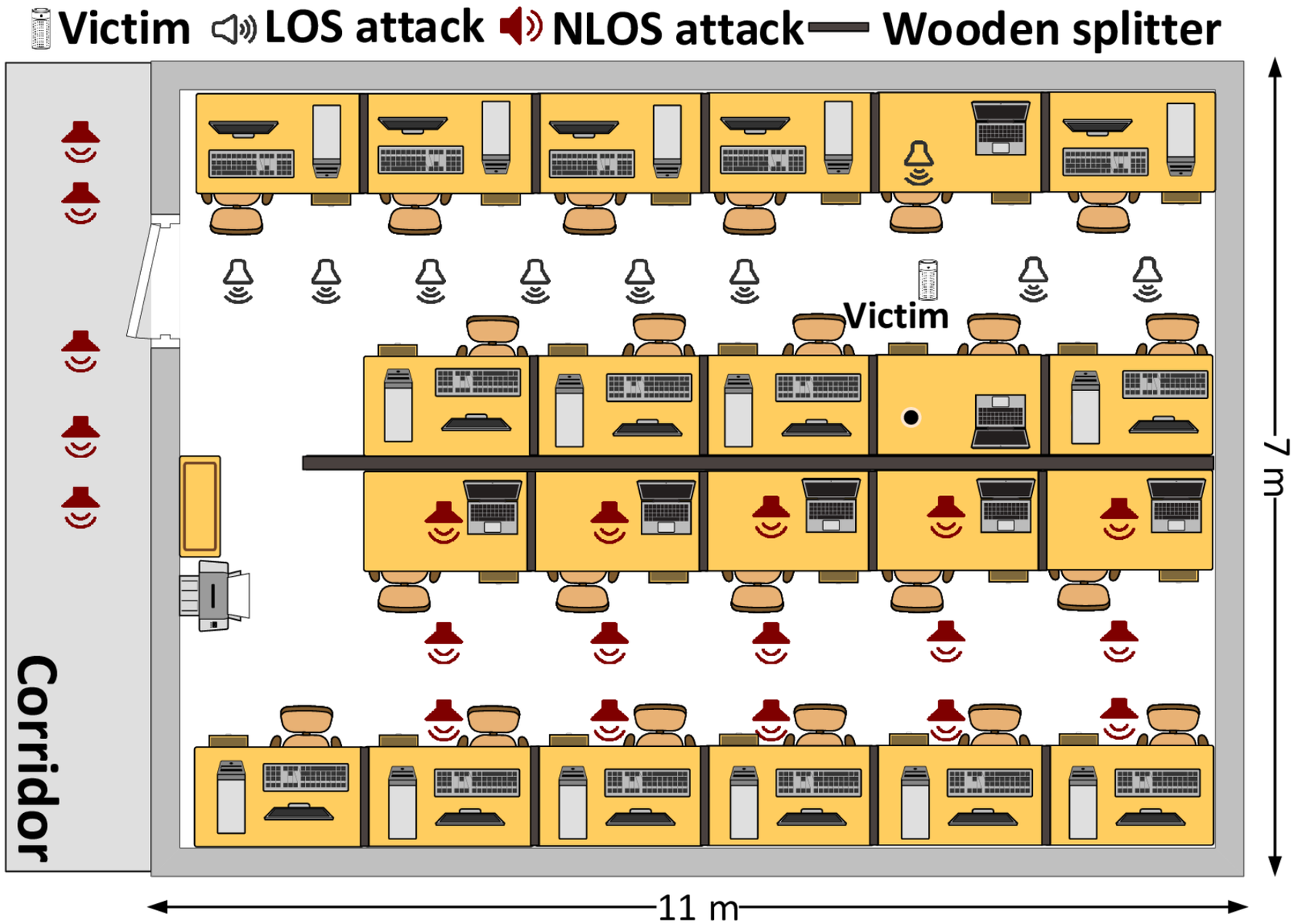
Original:
“your son went to serve at a distant place and became a centurion”

Meta-Enha:
“open the door”

Meta-Qual:
“open the door”

Evaluation: Attack Successful Rate

- Attack Target: “DeepSpeech” (White-Box)

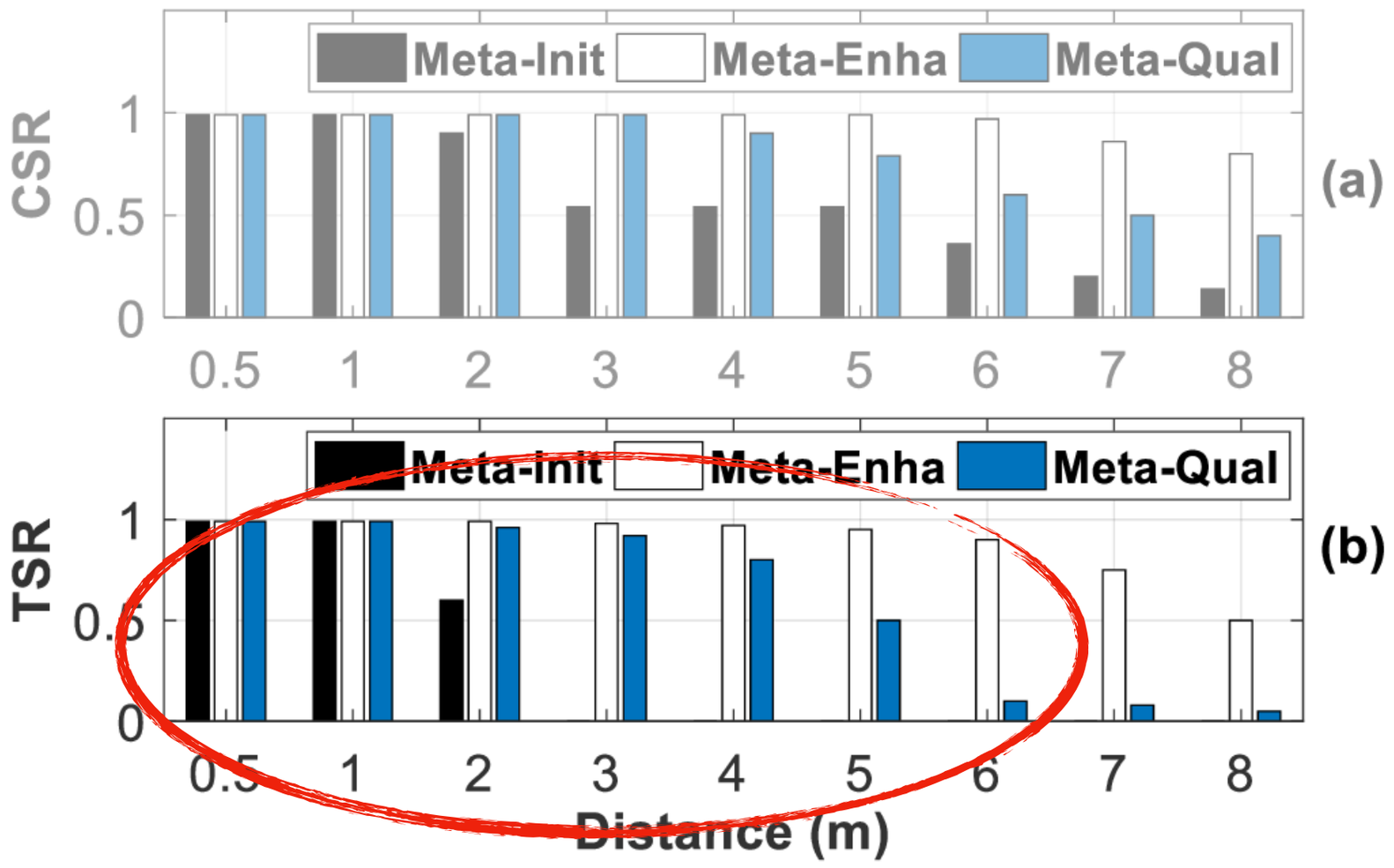


A multi-path prevalent office

Evaluation: Attack Successful Rate

- Line-of-Sight (LOS) Attack

Character Successful Rate

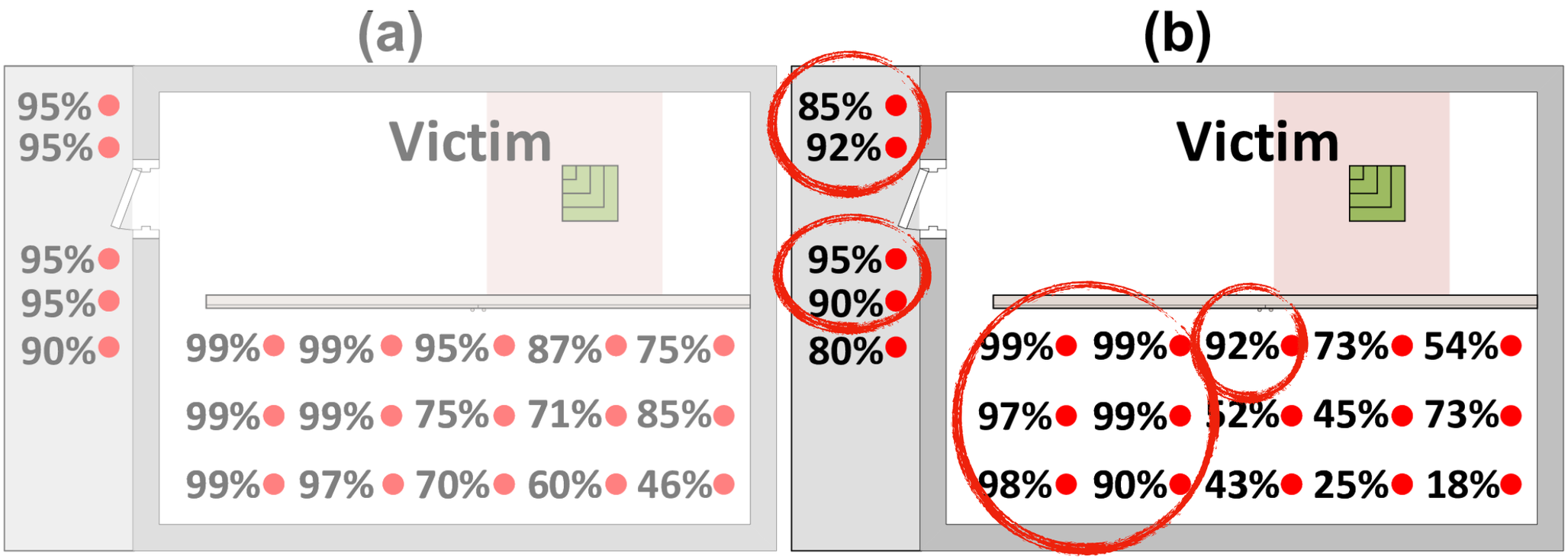


Transcript Successful Rate

Meta-Enha: > 90% attack successful rate

Evaluation: Attack Successful Rate

- No-Line-of-Sight (NLOS) Attack



Character Successful Rate

Transcript Successful Rate

Meta-Enha: over 85% attack successful rate across 11/20 NLOS location!

Conclusion

1. Investigate over-the-air audio adversarial attacks systematically.
2. Propose a “generate-and-clean” two-phase design and improve the audio quality.
3. Develop a prototype and conduct extensive evaluations.

Visit acoustic-metamorph-system.github.io for more information!