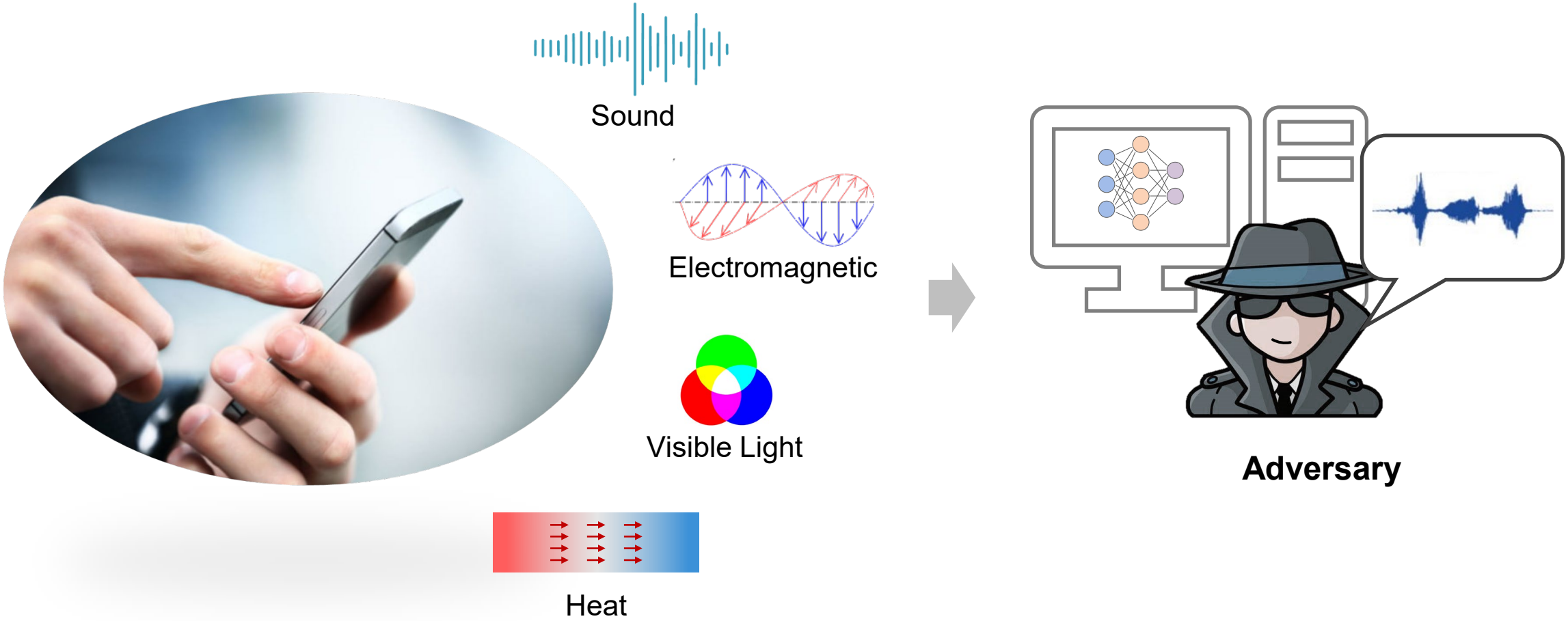
浙江大學
ZHEJIANG UNIVERSITY

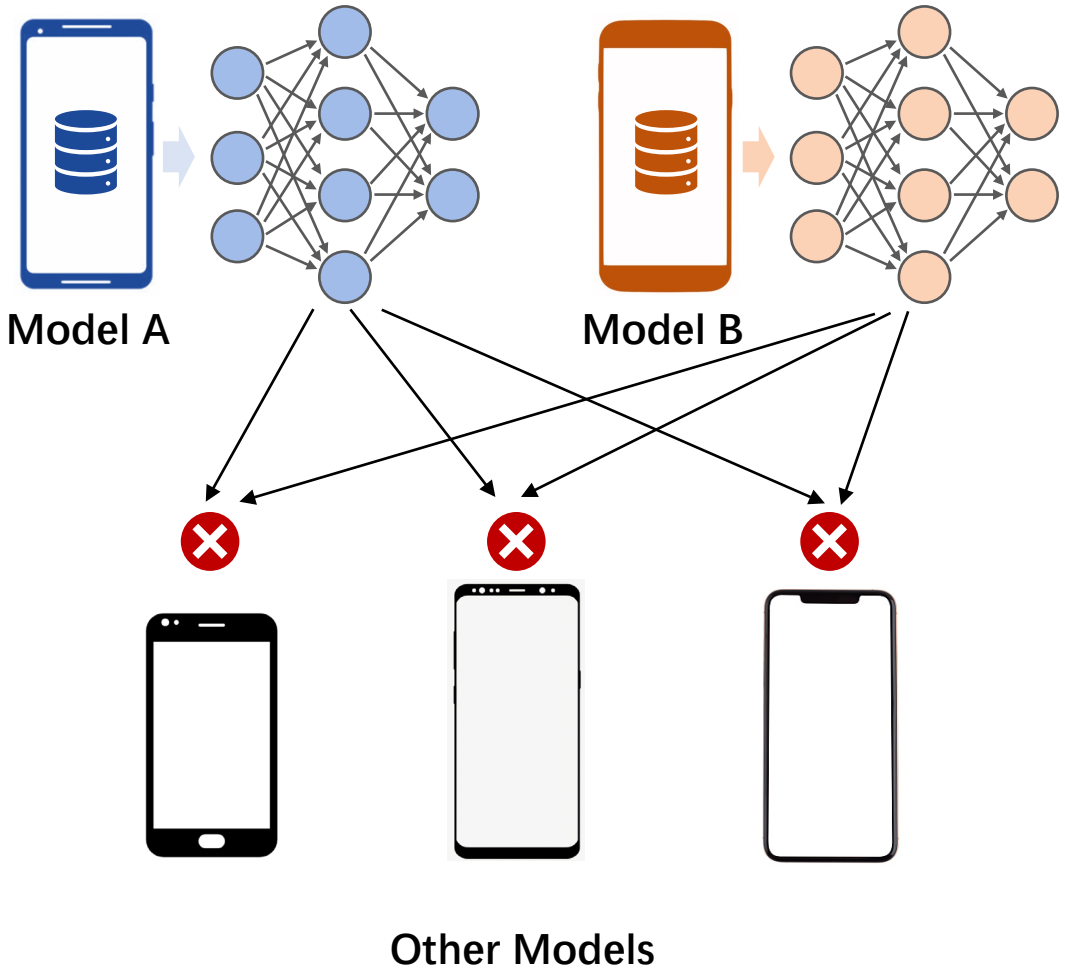
InertiEAR: Automatic and Device-independent IMU-based Eavesdropping on Smartphones

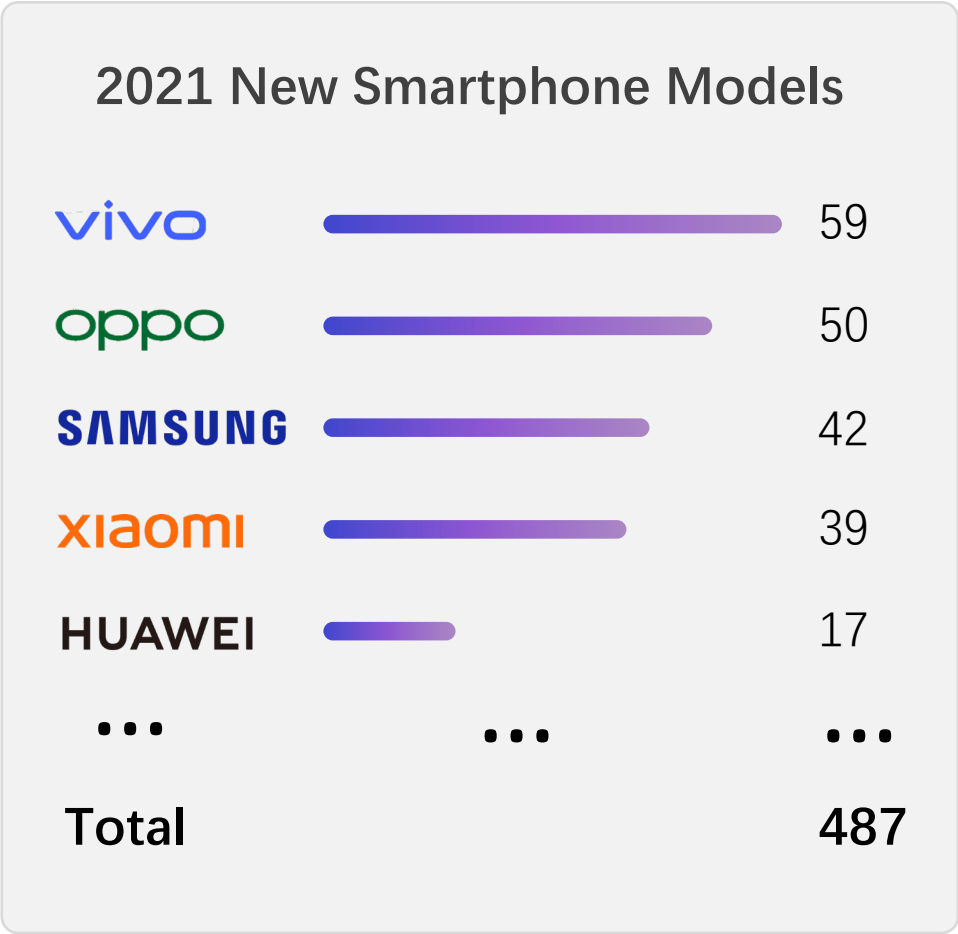
Ming Gao, Yajie Liu, Yike Chen, Yimin Li, Zhongjie Ba, Xian Xu, Jinsong Han

Side-Channel Eavesdropping

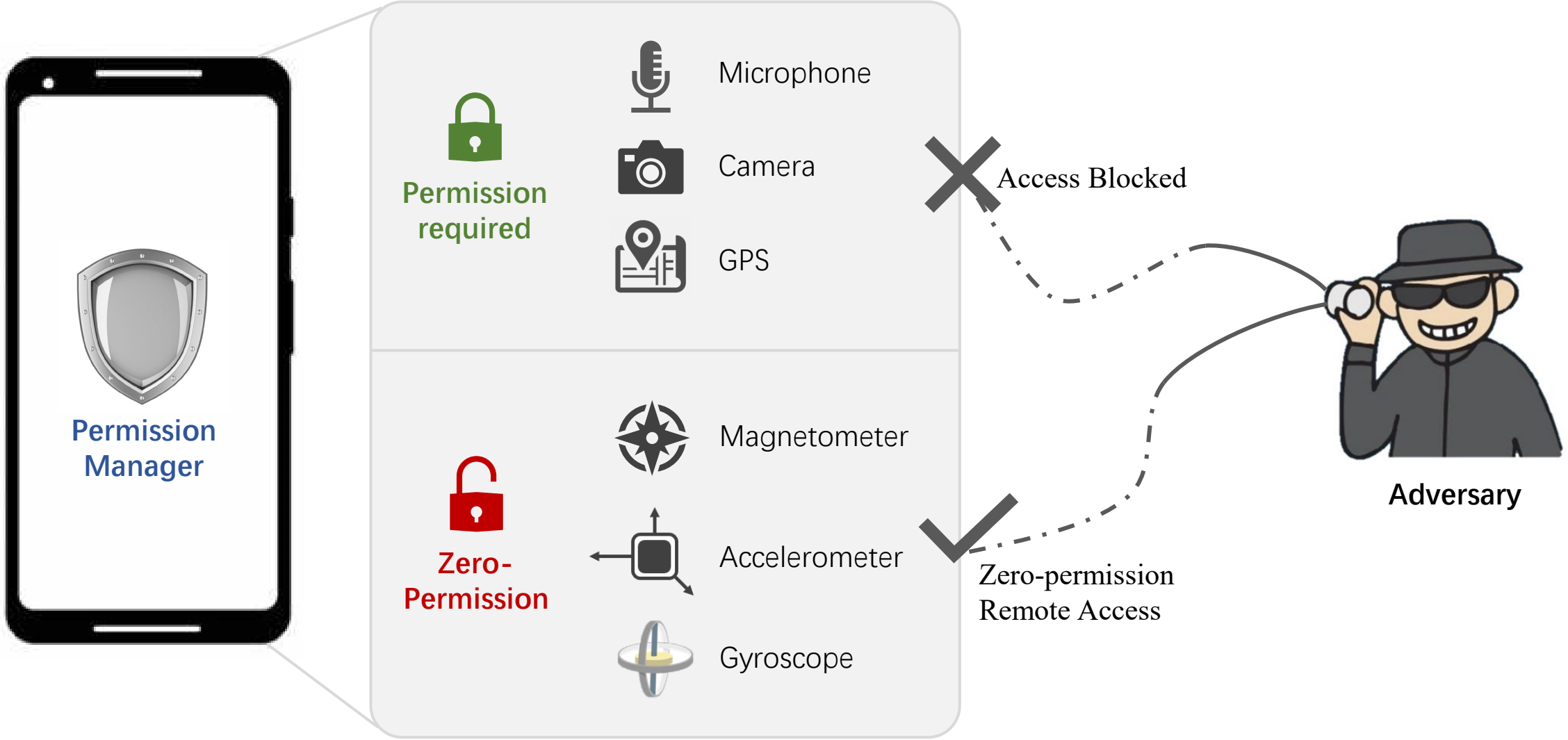


Smartphone Models

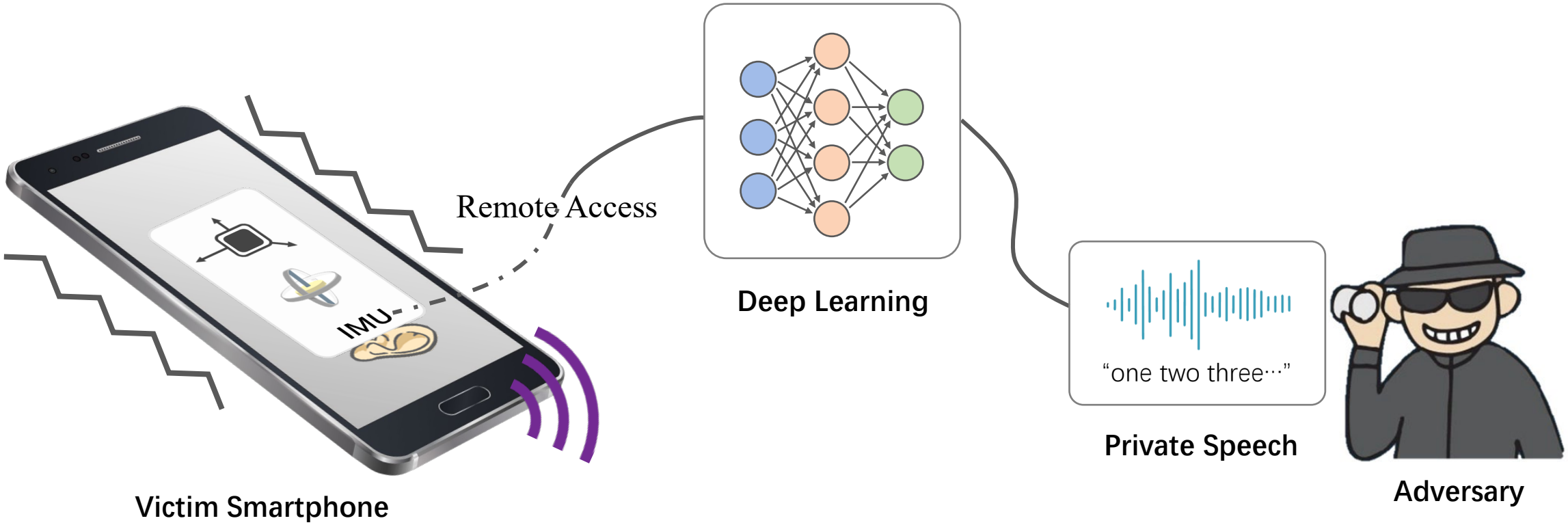




Smartphone Sensors



Motion Sensor Threat to Speech Privacy

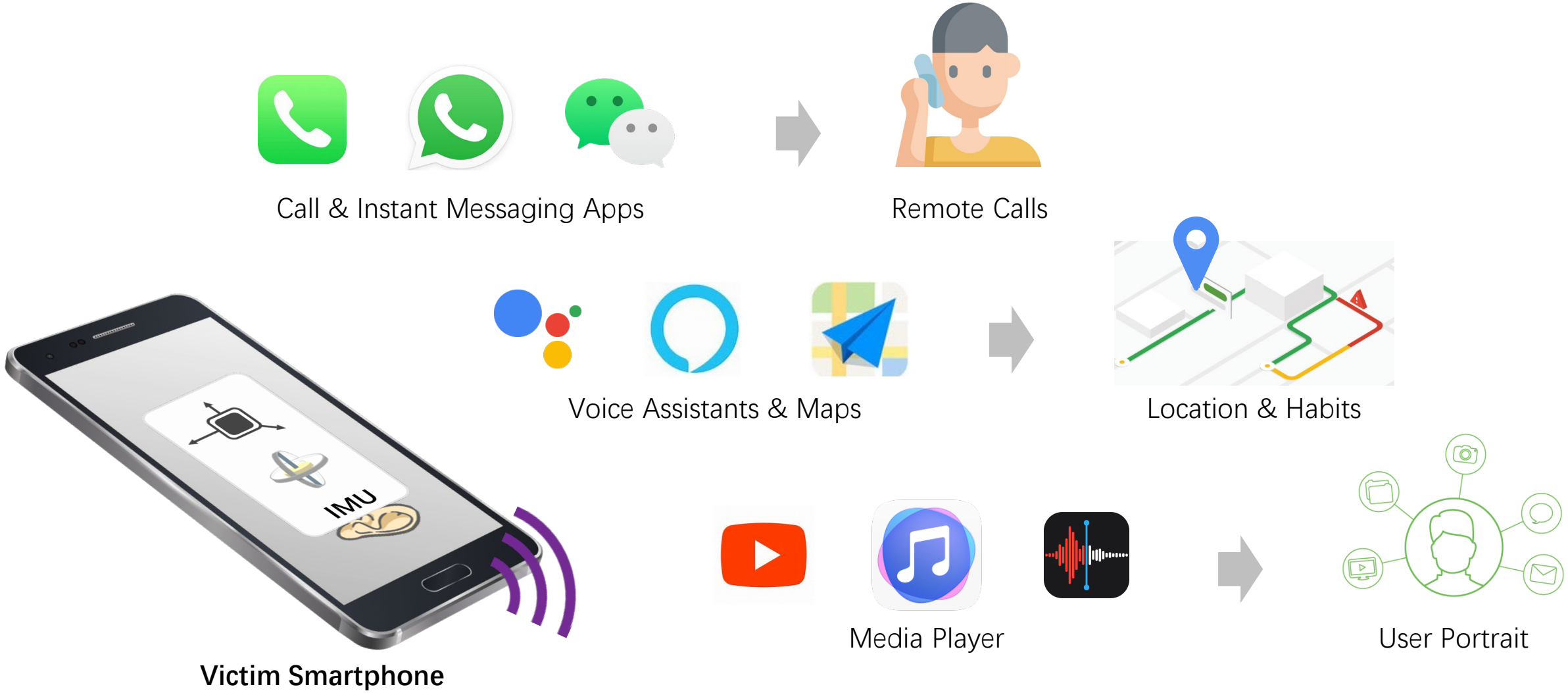


[1] Y. Michalevsky, D. Boneh, and G. Nakibly, "Gyrophone: Recognizing speech from gyroscope signals," in USENIX Security Symposium, 2014.

[2] Z. Ba, T. Zheng, X. Zhang, Z. Qin, B. Li, X. Liu, and K. Ren, "Learning-based practical smartphone eavesdropping with built-in accelerometer," in NDSS, 2020.

[3] S. A. Anand, C. Wang, J. Liu, N. Saxena, and Y. Chen, "Spearphone: A lightweight speech privacy exploit via accelerometer-sensed reverberations from smartphone loudspeakers," in ACM WiSec, 2021.

Motion Sensor Threat to Speech Privacy

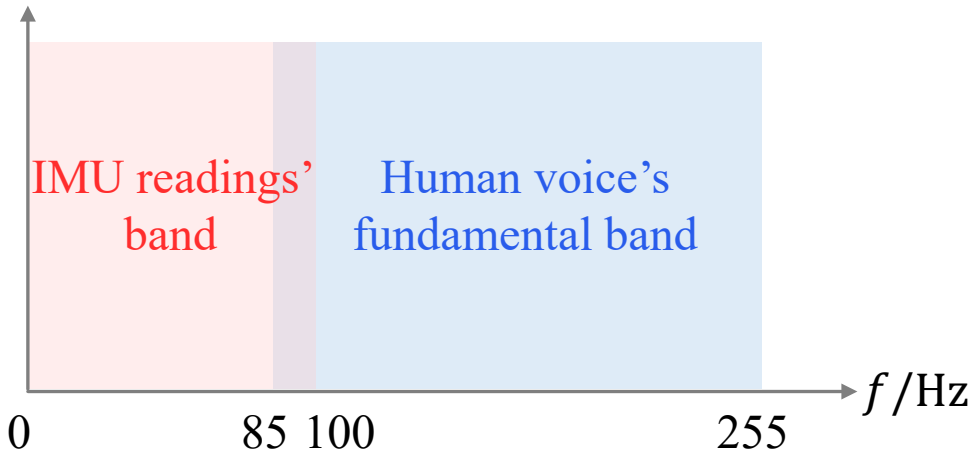


Countermeasure: Lower the volume

Victim Smartphone

Volume Setting		Bottom Speaker		
		20%	60%	100%
Acc	a_x	0.69	2.21	3.07
	a_y	4.24	5.49	5.88
	a_z	4.84	5.07	5.19
Gyro	ω_x	-7.66	-4.28	-6.18
	ω_y	-7.01	-5.04	-5.63
	ω_z	-6.70	-6.42	-5.56

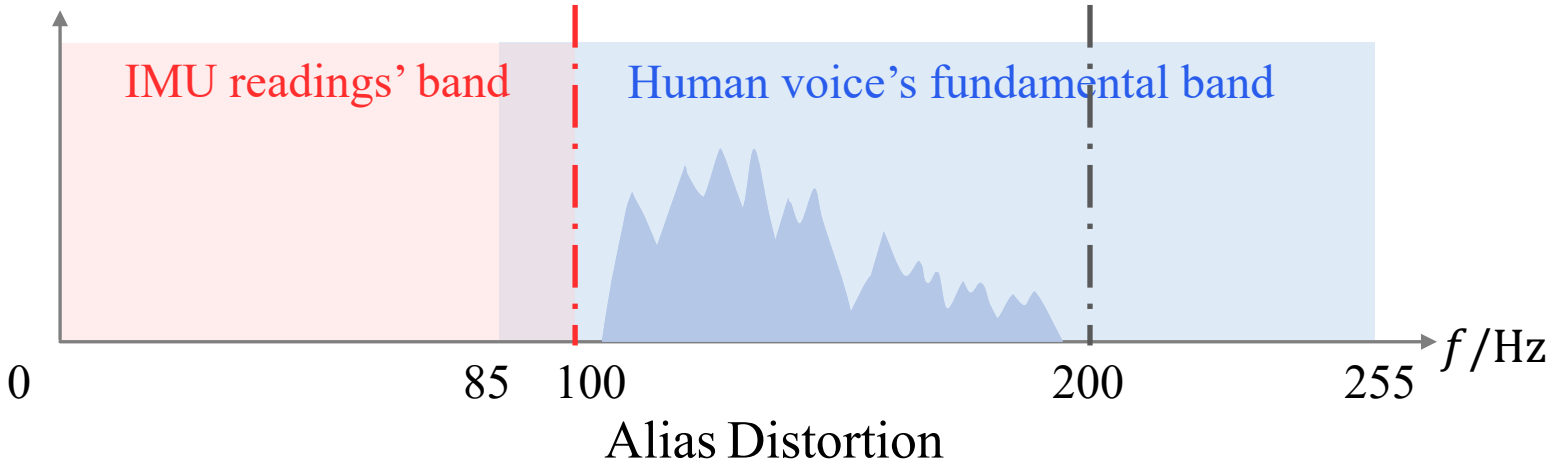
Countermeasure: Sampling rate limitation



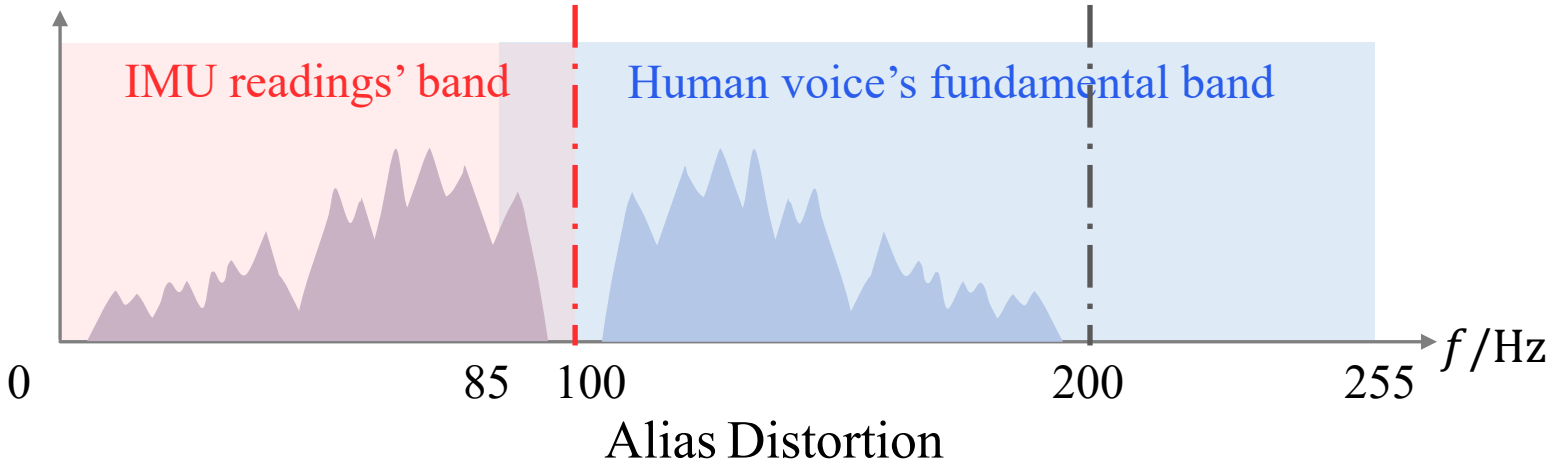
Android 12 requires $Fs \leq 200Hz$ in inertial sensors!

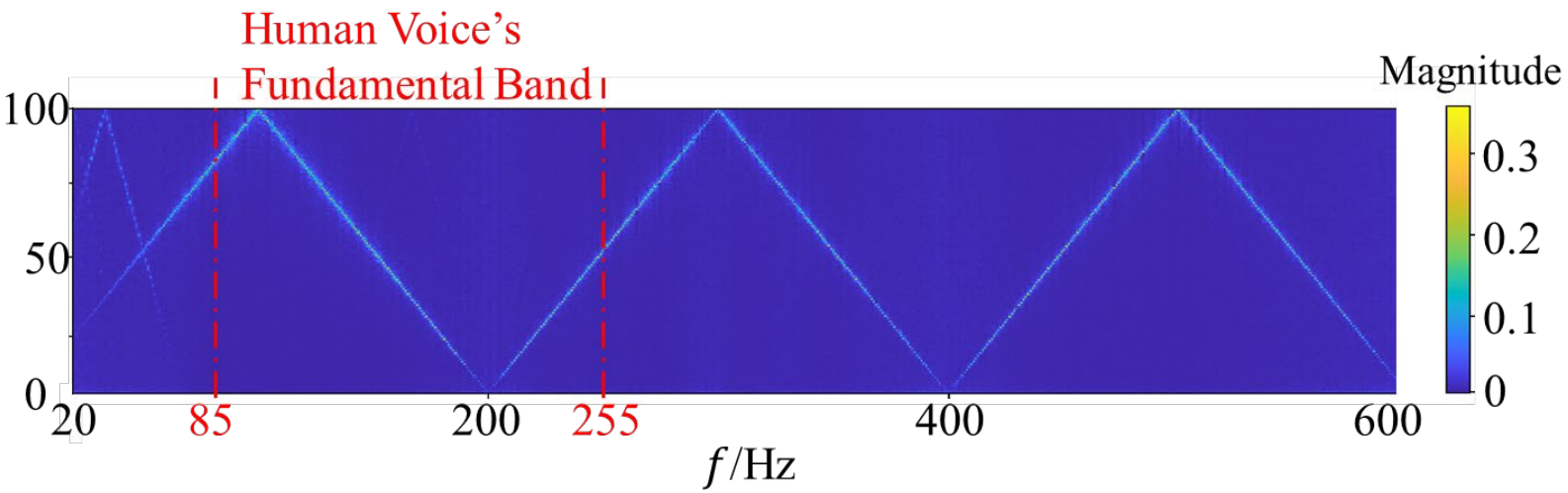
[1] Android for Developers, "Behavior changes: Apps targeting android 12," https://developer.android.com/about/versions/12/behavior-changes-12#motion-sensor-rate-limiting, 2021.

Our Observations



Our Observations



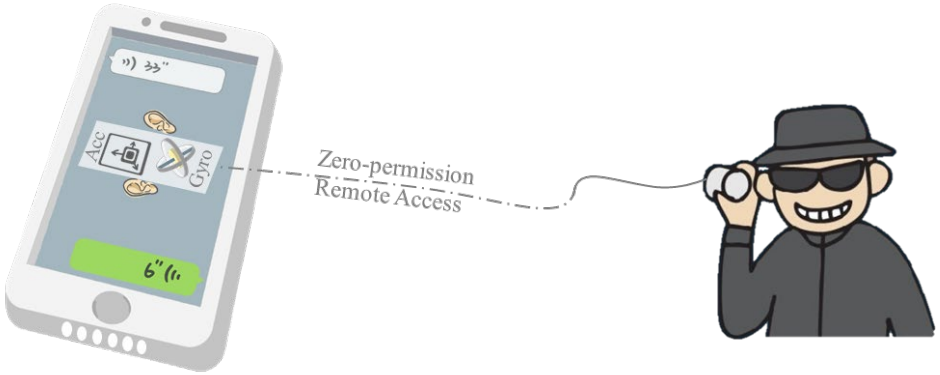


Smartphones' IMUs can respond to audio signals of up to 6 kHz

Our Vision: InertiEAR

IMU-based eavesdropping on smartphones:

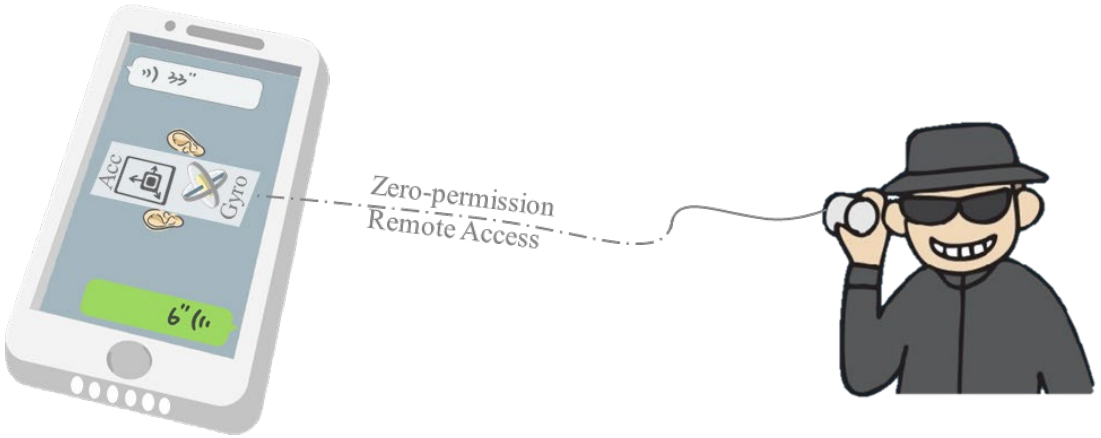
- ✓ Zero-permission
- ✓ Sampling rates within 200 Hz
- ✓ Automatic
- ✓ Device-independent



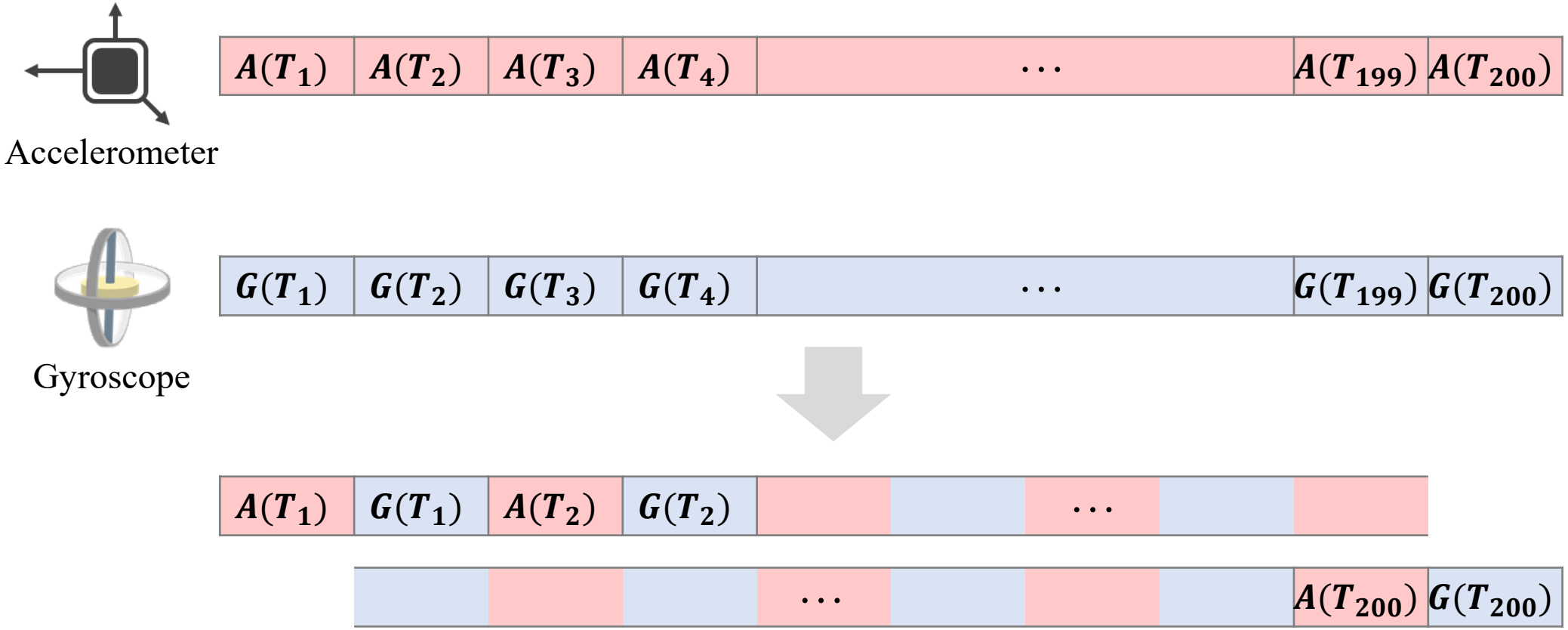
Challenges

1.How to eavesdrop accurately using IMUs whose sampling rates are limited within 200Hz?

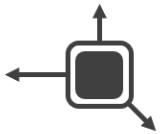
2.How to achieve device-independent eavesdropping?



Sensor Fusion



Speaker-to-IMU Channel



Maximum volume

Audio signals

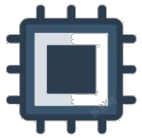
$$A(t) = k_i \cdot M \cdot H_a \cdot S_m(t) + N_a$$

$$G(t) = k_i \cdot M \cdot H_g \cdot S_m(t) + N_g$$

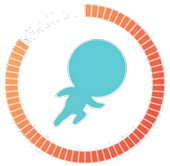
Volume level set by user

Gain coefficients

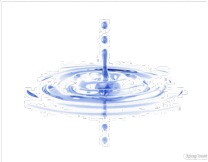
Channel noises



Intrinsic noise

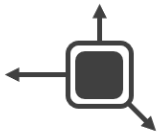


Motion interference



Harmonic

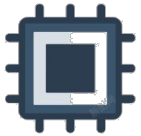
Speaker-to-IMU Channel



$$A(t) = k_l \cdot M \cdot H_a \cdot S_m(t) + N_a$$

$$G(t) = k_l \cdot M \cdot H_g \cdot S_m(t) + N_g$$

Channel noises

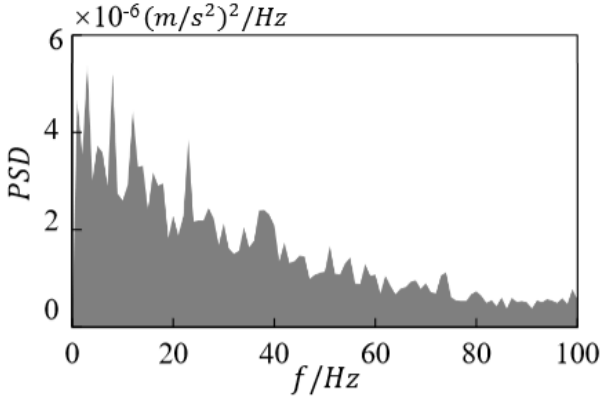


Intrinsic noise

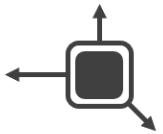
Motion interference

Harmonic

DC bias + white noise



Speaker-to-IMU Channel

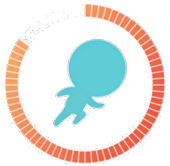


$$A(t) = k_l \cdot M \cdot H_a \cdot S_m(t) + N_a$$

$$G(t) = k_l \cdot M \cdot H_g \cdot S_m(t) + N_g$$

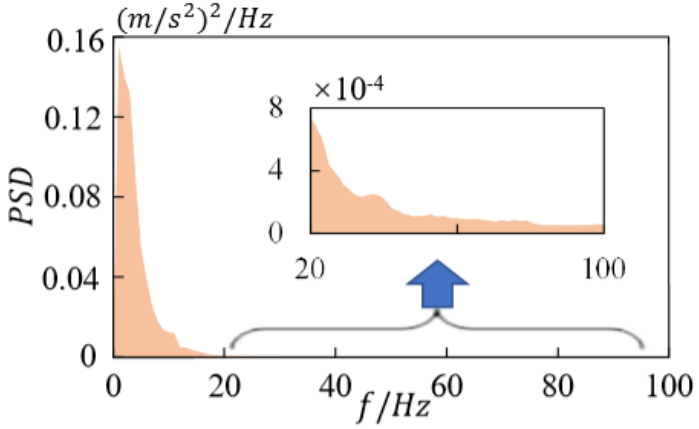
Channel noises

Intrinsic noise

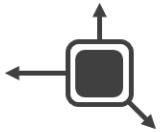


Motion interference

Harmonic



Speaker-to-IMU Channel



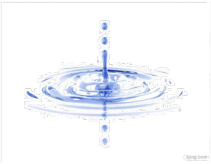
$$A(t) = k_l \cdot M \cdot H_a \cdot S_m(t) + N_a$$

$$G(t) = k_l \cdot M \cdot H_g \cdot S_m(t) + N_g$$

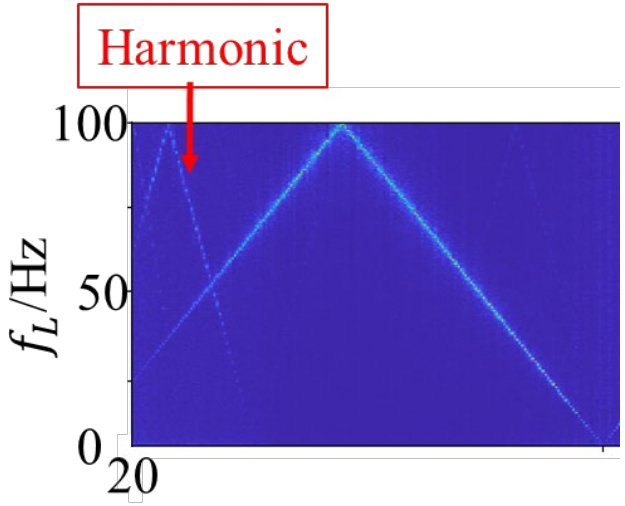
Channel noises

Intrinsic noise

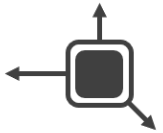
Motion interference



Harmonic



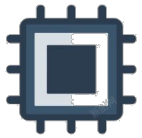
Speaker-to-IMU Channel



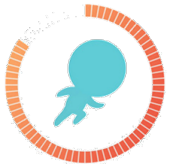
$$A(t) = k_l \cdot M \cdot H_a \cdot S_m(t) + N_a$$

$$G(t) = k_l \cdot M \cdot H_g \cdot S_m(t) + N_g$$

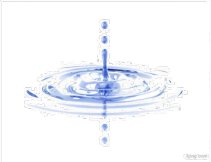
Channel noises



Intrinsic noise

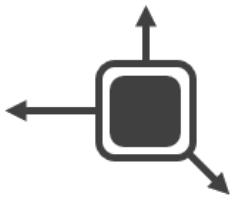


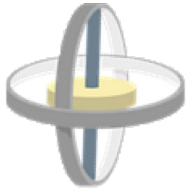
Motion interference

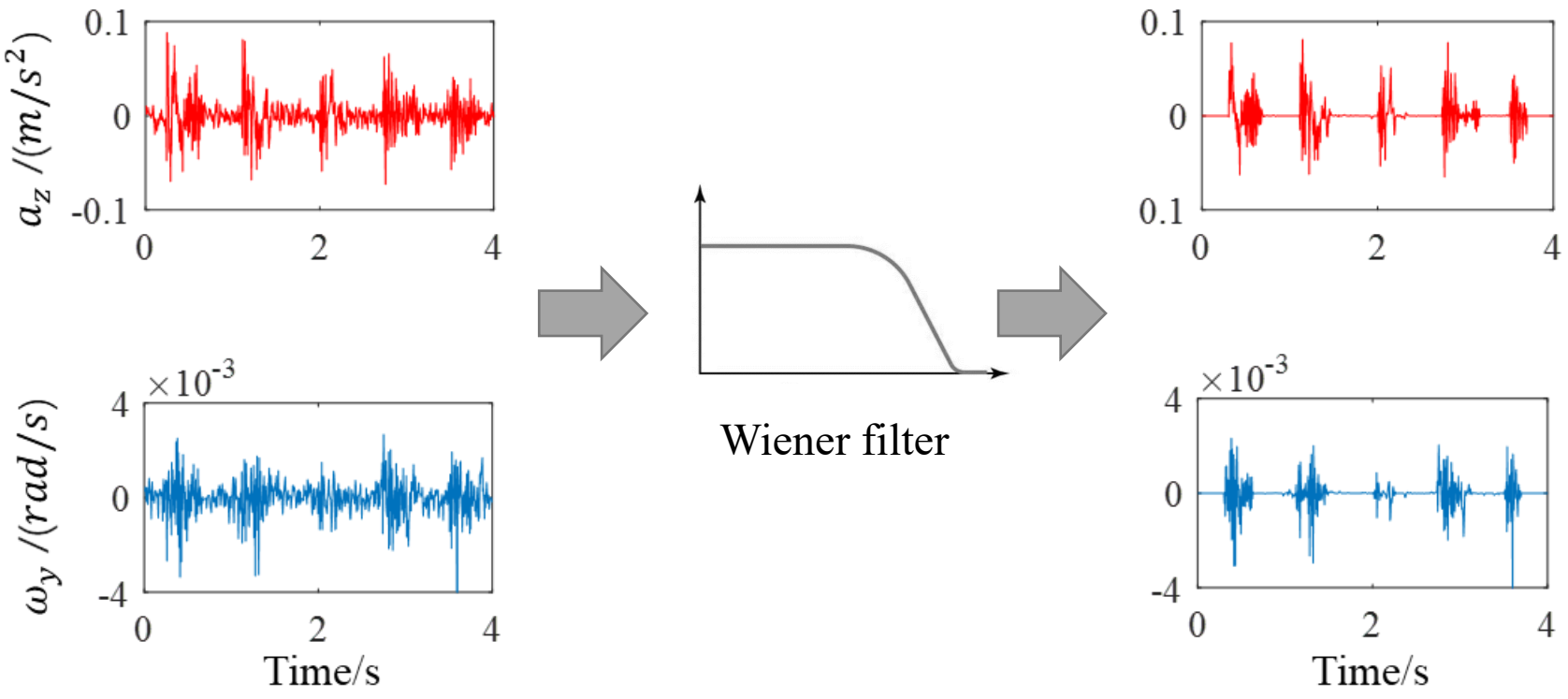


Harmonic

Noise Elimination

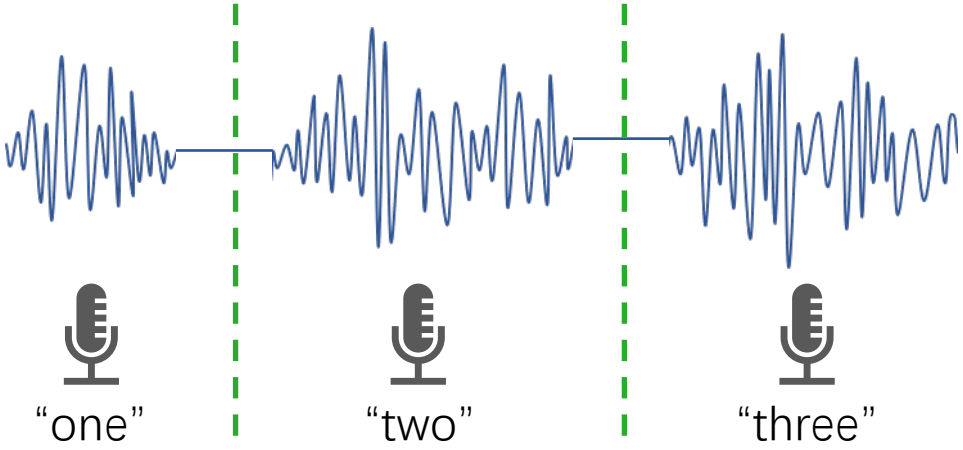




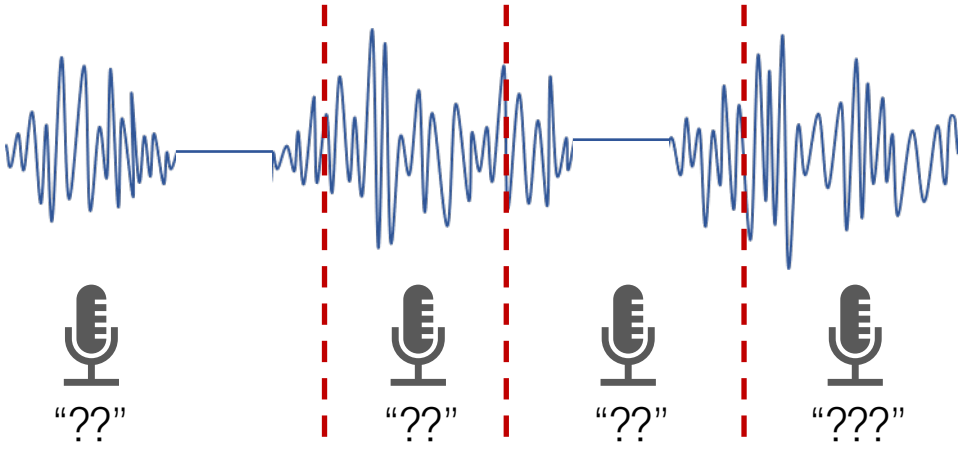


Automatic Segmentation

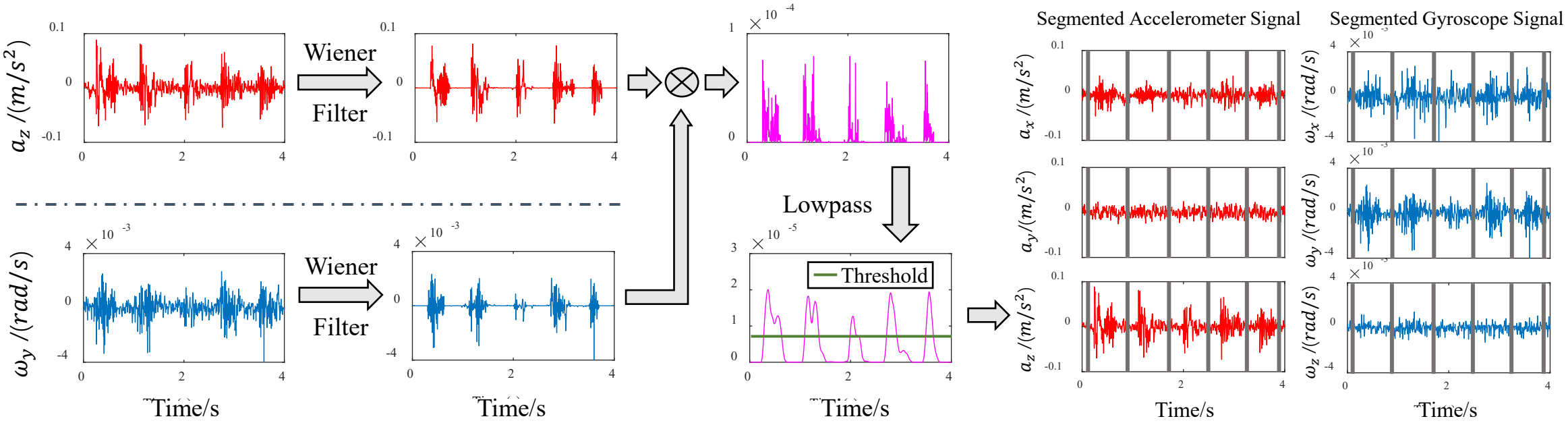
Precise Segmentation



Imprecise Segmentation



Automatic Segmentation



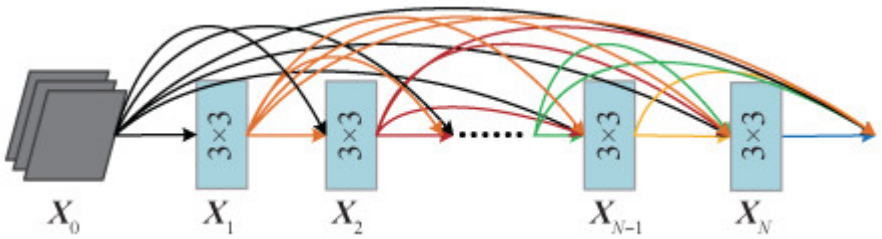
$$a_z(t) = k_a sin(2\pi f_L t) + n_{waz}(t) + m_{az}(t) + n_{haz}(t)$$
$$g_x(t) = k_g sin(2\pi f_L t) + n_{wgx}(t) + m_{gx}(t),$$

$$a_z(t) \times g_x(t) = \boxed{\frac{k_a k_g}{2}}$$

DC

Speech Recognition

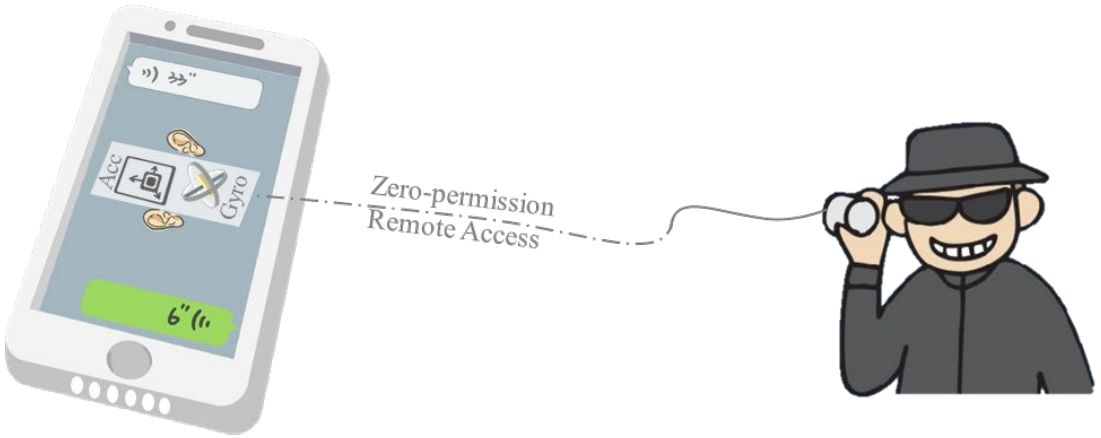
- Data processing:
  244 × 244 gray spectrogram-images
- DenseNet:
  - A dense connection between all the previous layers to the layers behind
  - Less computational cost



Challenges

1.How to eavesdrop accurately using IMUs whose sampling rates are limited within 200Hz?

2.How to achieve device-independent eavesdropping?



Hardware Diversity

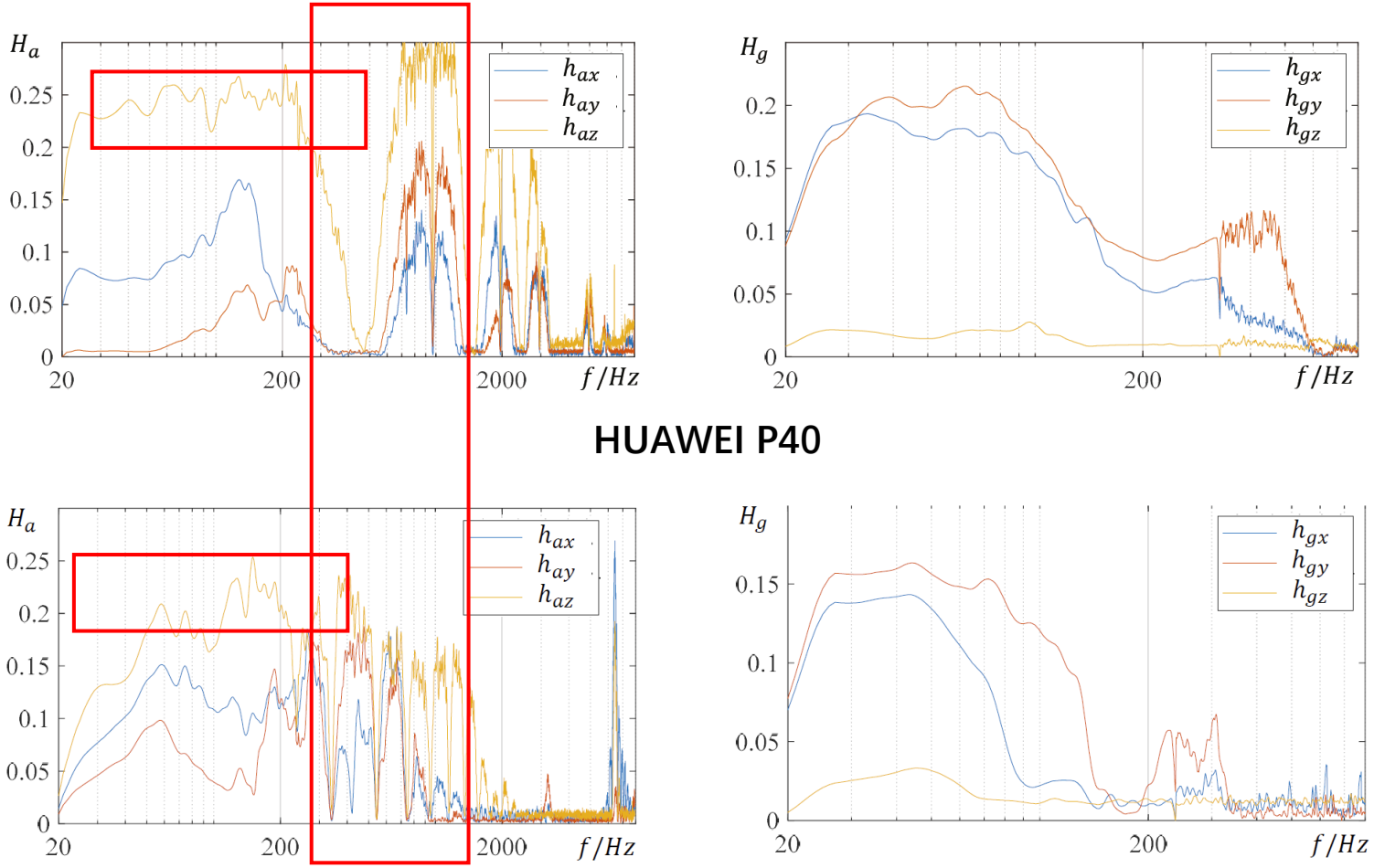
$$A(t) = k_l \cdot M \cdot H_a \cdot S_m(t) + N_a$$

Hardware Diversity

$$A(t) = k_l \cdot M \cdot H_a \cdot S_m(t) + N_a$$

Axial Energy Rate

Frequency Response



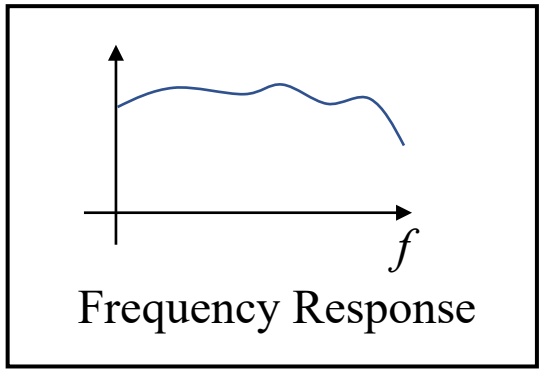
HUAWEI P40

Samsung Galaxy S8

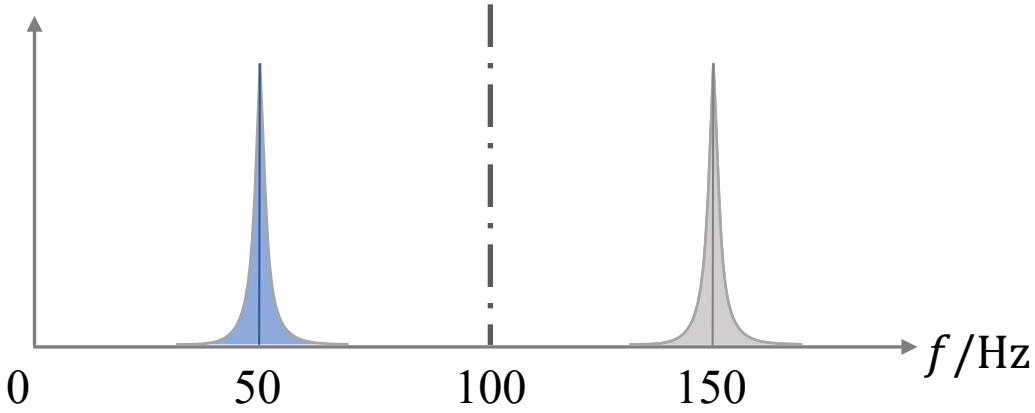
Hardware Diversity

$$A(t) = k_l \cdot M \cdot H_a \cdot S_m(t) + N_a$$

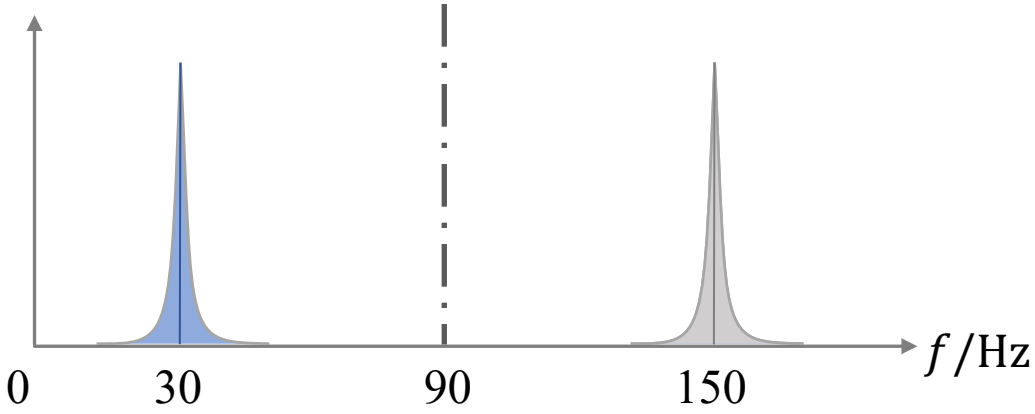
Axial Energy Rate



Sampling Rate



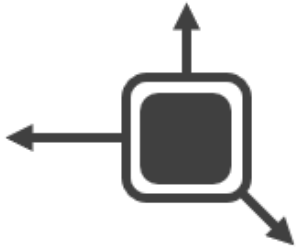
Fs = 200 Hz



Fs = 180 Hz

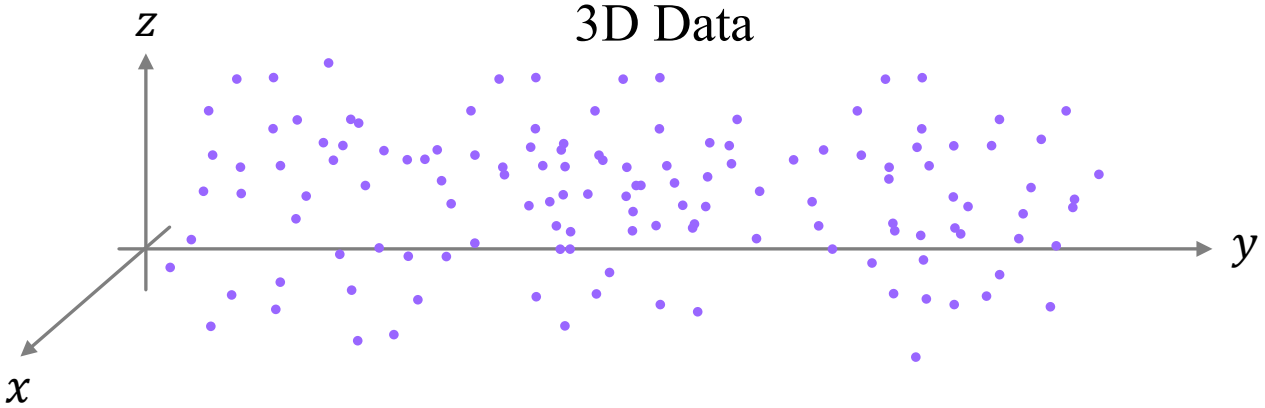
Solutions: Dimension Reduction

$$A(t) = k_l \cdot M \cdot H_a \cdot S_m(t) + N_a$$

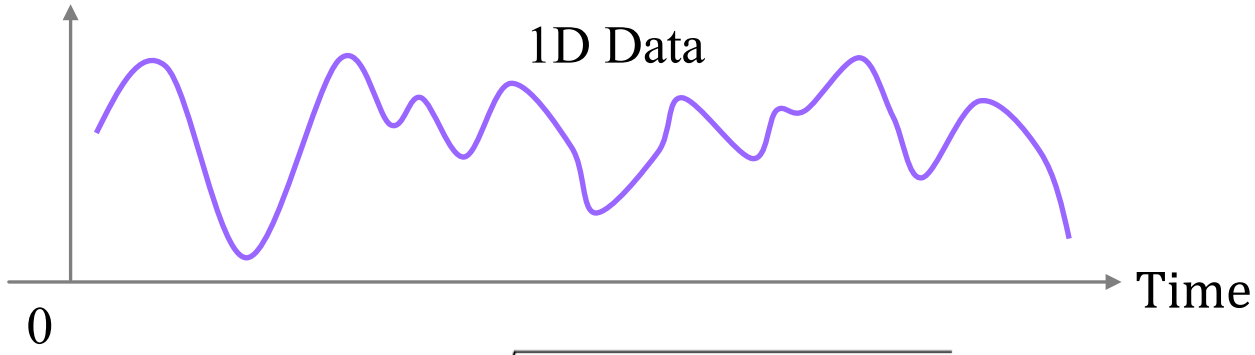


Axial Energy Rate

Frequency Response



Dimension Reduction



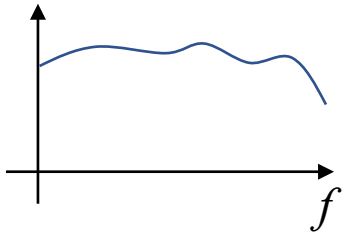
$$||\boldsymbol{H_i}||(t) = \sqrt{h_{ix}^2(t) + h_{iy}^2(t) + h_{iz}^2(t)},$$

$$A^{\dagger}(t) = sign(a_{max}(t))||\boldsymbol{A}||(t),$$

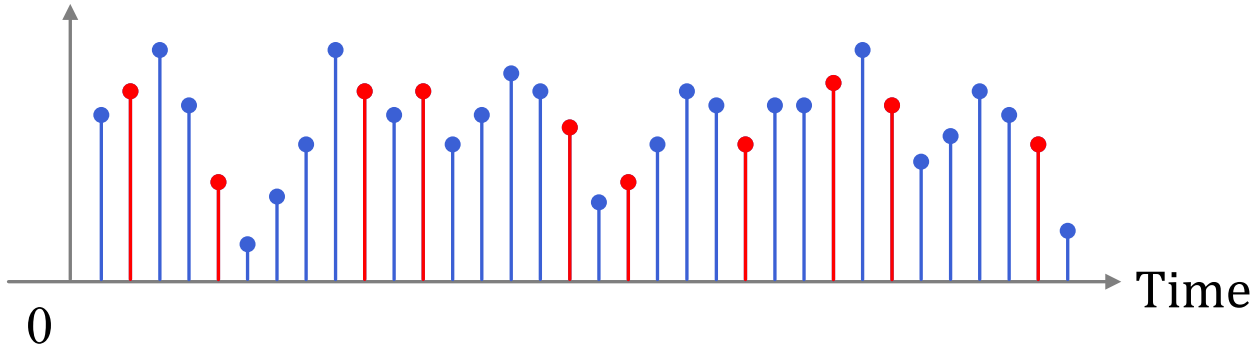
Solutions: High-frequency Suppression

$$A(t) = k_l \cdot M \cdot H_a \cdot S_m(t) + N_a$$

Axial Energy Rate



Frequency Response

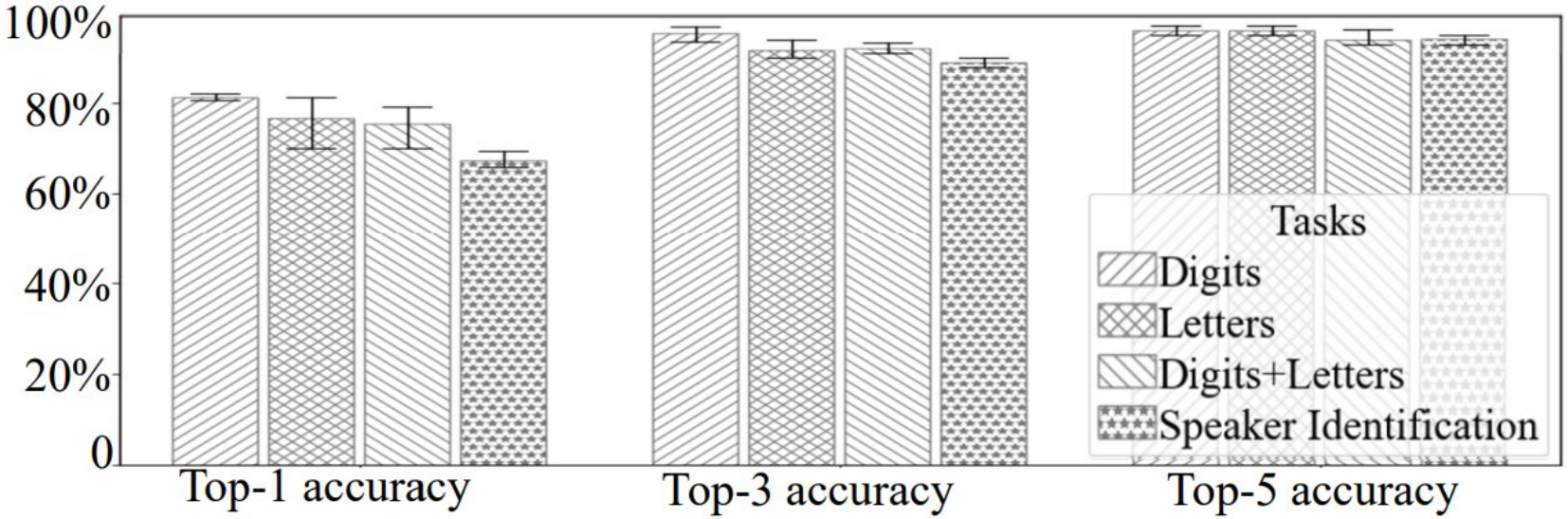


Random Down Sampling

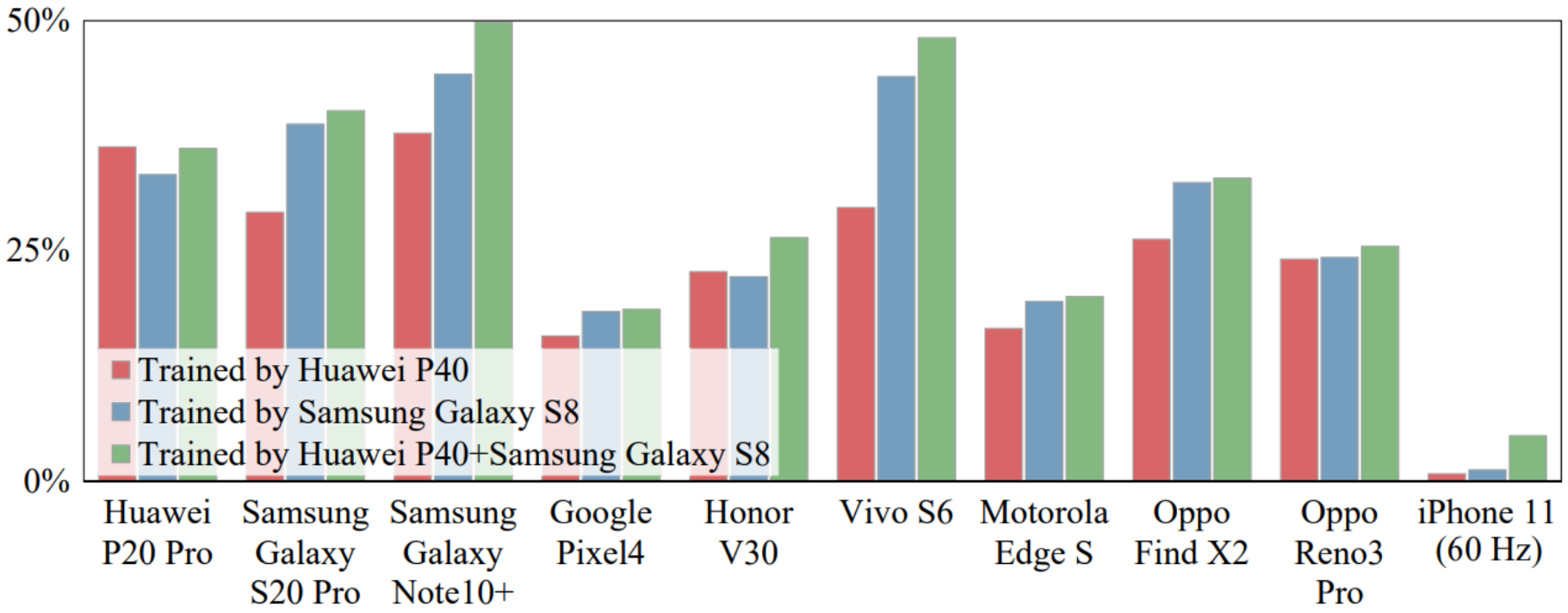
$$SNR = -20log_{10}(2\pi f \times rms(T_a))$$

Evaluation


Harmony OS


100%
80%
60%
40%
20%
0
Top-1 accuracy
Top-3 accuracy
Top-5 accuracy
Tasks
Digits
Letters
Digits+Letters
Speaker Identification

Cross-device Performance



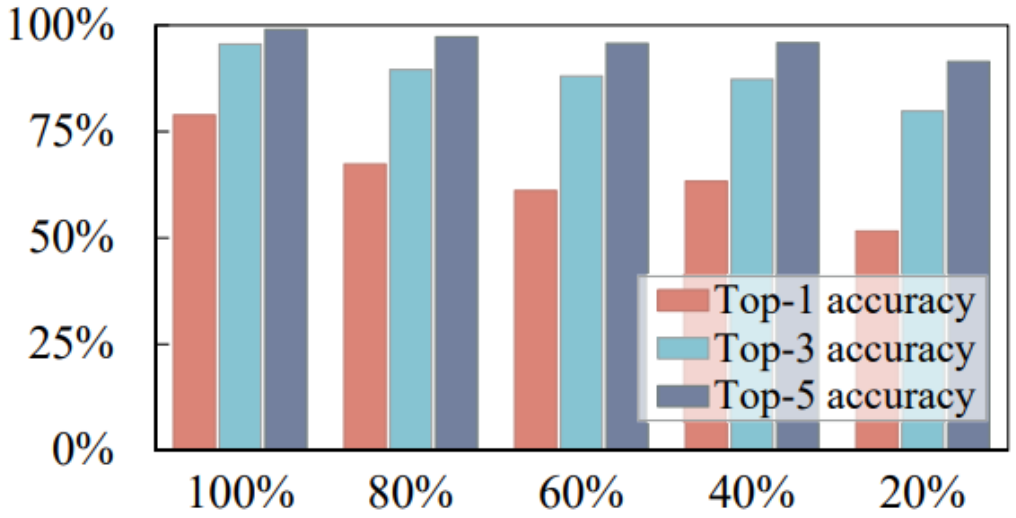
Cross-model accuracy

InertiEAR results:
Up to 49.8%

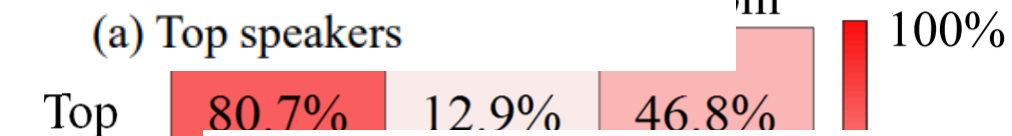
v.s.

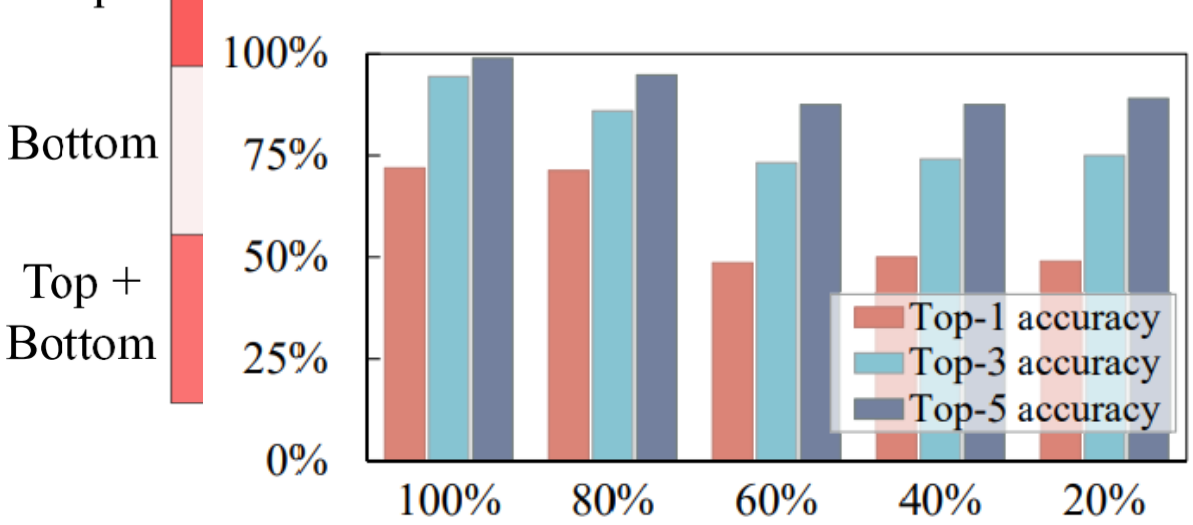
Previous SOTA results:
At most 26%

Impac



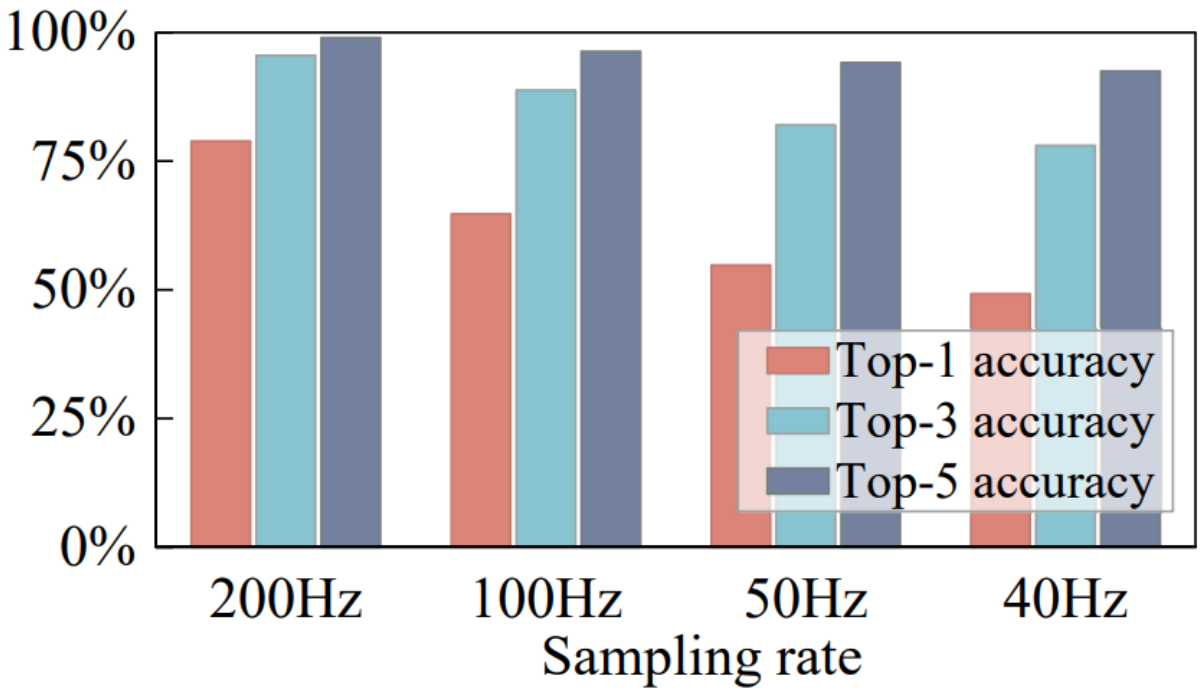
(a) Top speakers





(b) Bottom speakers

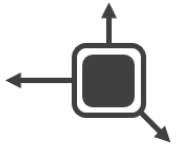
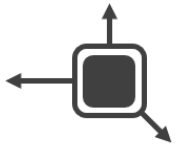
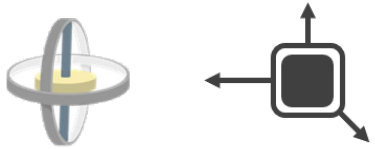
Impact of Sampling Rate



Even the limitation of 40 Hz sampling rate is still at risk!

Limitations on sampling rates barely work!

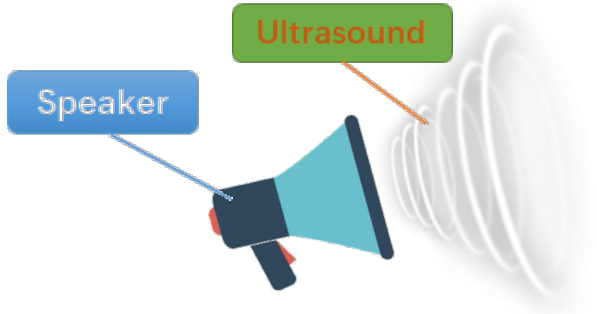
Comparison with SOTA attacks

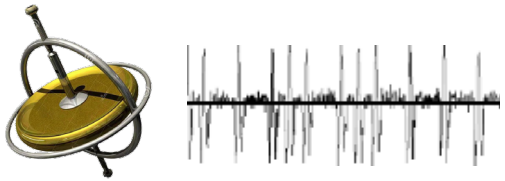
	Gyrophone [USENIX Security' 14]	AccelEve [NDSS'20]	Spearphone [WiSec'21]	InertiEAR
Sensor				
Sampling Rate	200 Hz	500 Hz	4 kHz	200 Hz
Speech Recognition	26%	78%	81%	78.8%
Segmentation	Manually	92%	82%	100%
Device Independence	Not learning-based	at most 26%	×	up to 49.8%

Defense

Existing methods:

- Sampling rate limitation
- Filter
- Damping and isolating





Our suggestion:

- ➢ For users:
  - ✓ Resonant noise
- ➢ For manufacturers:
  - ✓ Oversampling

Conclusions

- ✓ We revisit the threat of IMU-based eavesdropping and realize a side channel attack InertiEAR. It breaks the restriction on sampling rates.
- ✓ A mathematical model is proposed to expand its attack surface and promote its practicality.
- ✓ InertiEAR accomplishes a device-independent eavesdropping attack.