DolphinAttack: Inaudible Voice Commands

Guoming Zhang Chen Yan Xiaoyu Ji *

Tianchen Zhang Taimin Zhang Wenyuan Xu *

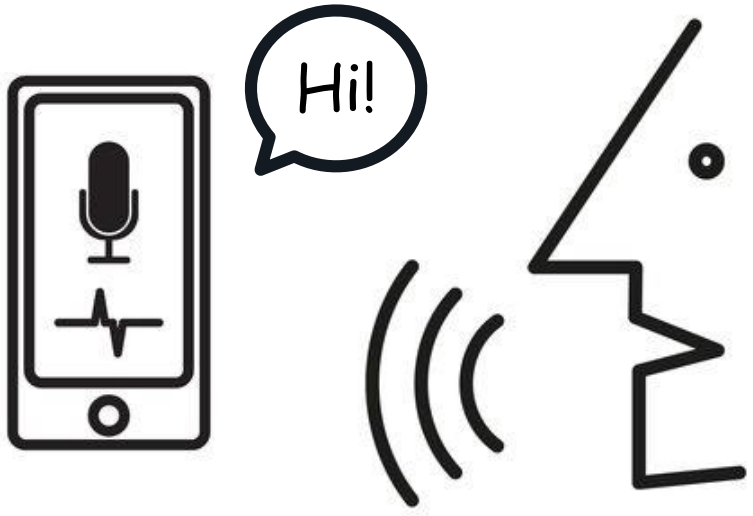
USSLAB, Zhejiang University

智能系统安全实验室
UBIQUITOUS SYSTEM SECURITY LAB.

Outline

1. Background of Voice Assistants
2. Design of DolphinAttack
3. Attack Scenarios
4. Evaluation
5. Defense & Responsible Disclosure

Voice becomes an increasingly important interface

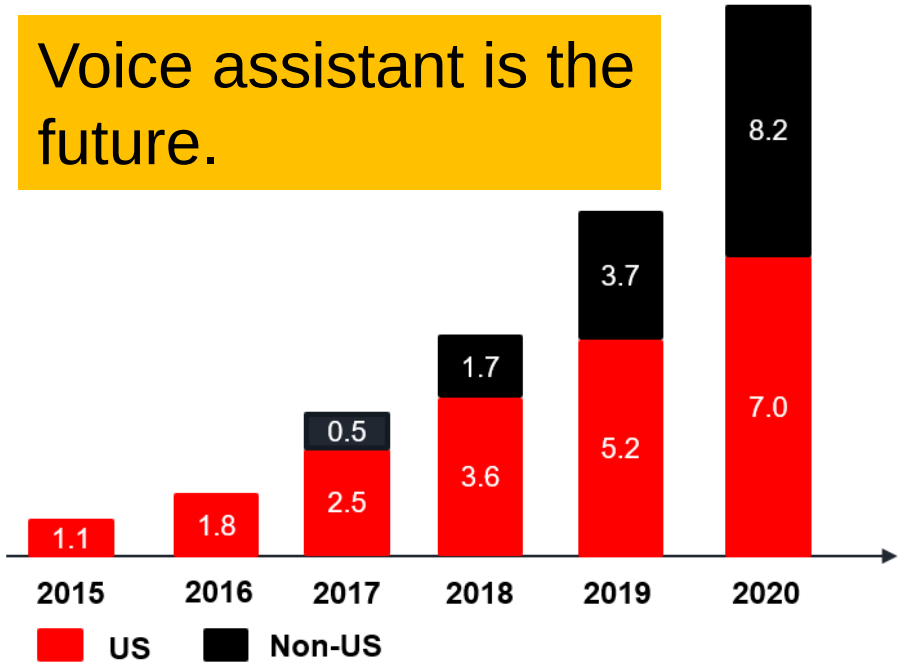


Siri Google Now Alexa Cortana S Voice Hi Voice

Digital Voice-activated Assistant Device Shipments Worldwide, US vs. Non-us, 2015-2020

millions

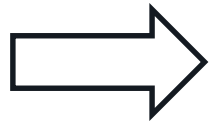
Voice assistant is the future.

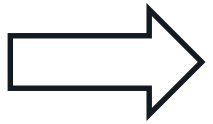


Source: Strategy Analytics as cited in press release, Aug,26,2016

How do voice assistants work?

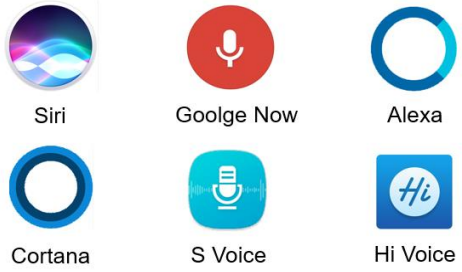
Sound *Electrical signal* *Command*





Voice Command Input

Speech Recognition



Command Execution

What can a malicious user achieve?

What's on my calendar today?

Sensitive information

Open evil.com

Malicious website

Tell my wife I love her

Fake message

Send an email to my boss

Social engineering

Open the front door

Break-in

Buy something on Amazon

Lose money

Transfer $100 to Eve

Steal money

Call 1234567890

Eavesdrop

Facetime Eve

Spy

Drive me to Austin

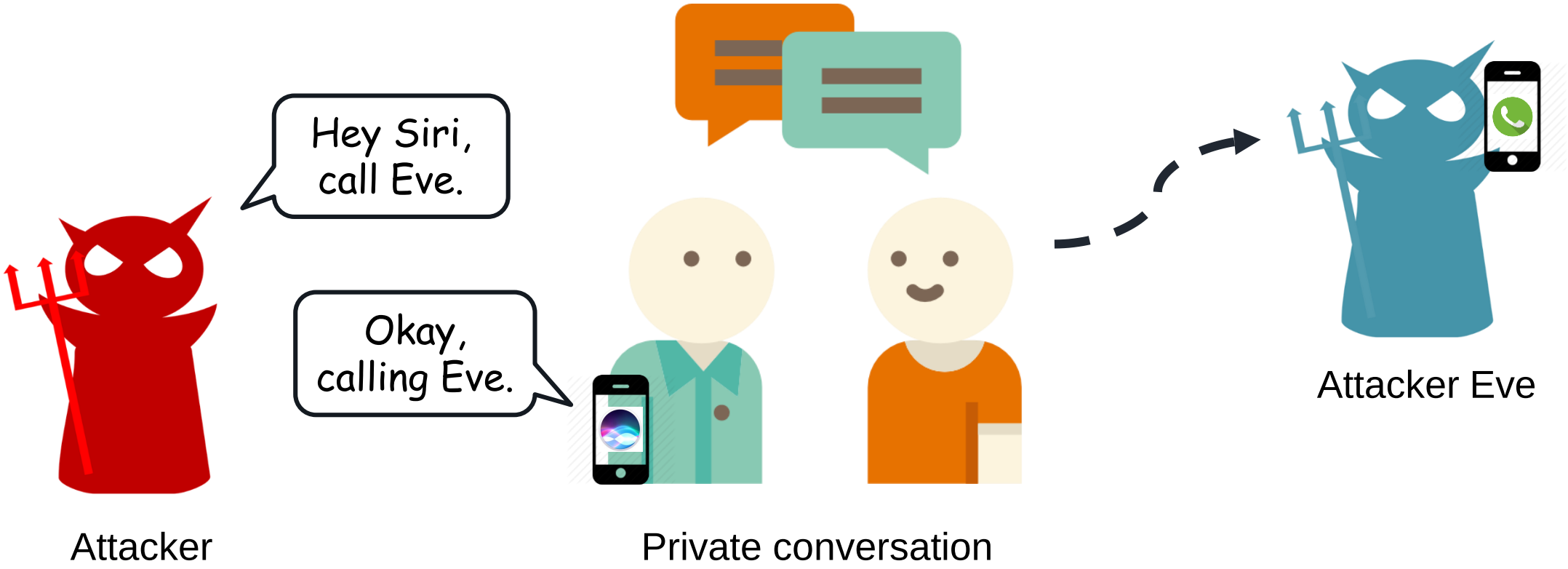
Mislead

Attack Scenario 1: fake online orders

Attacker

Amazon Echo

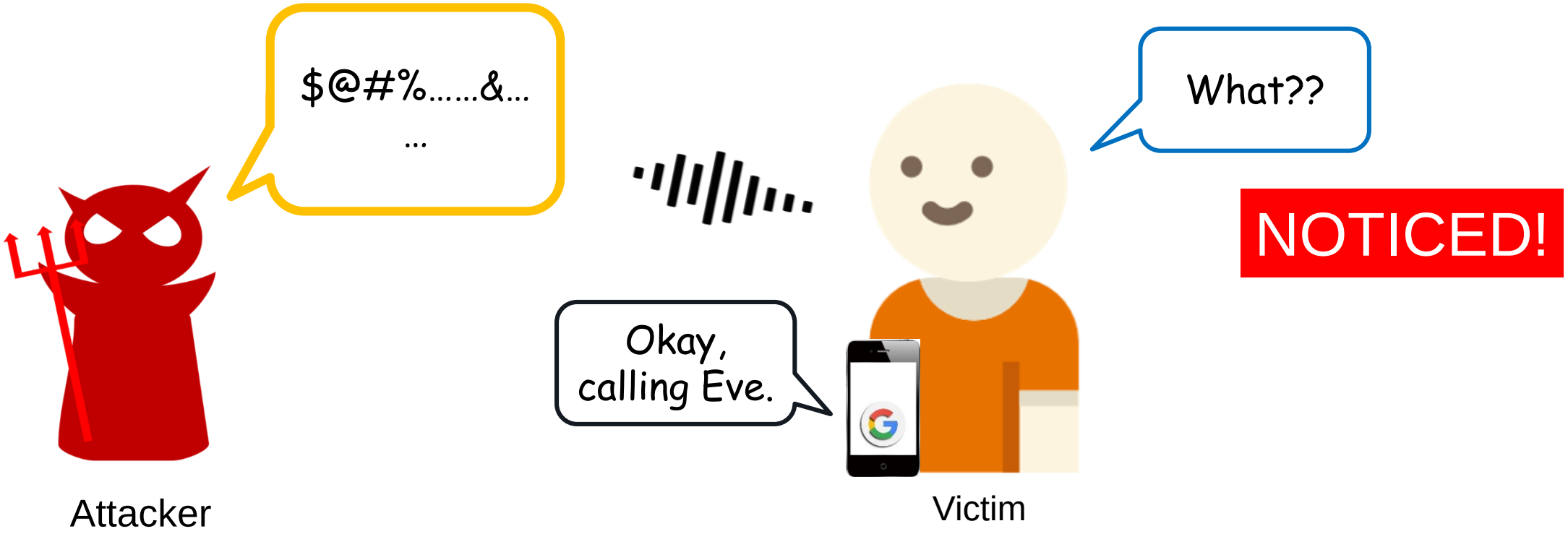
Attack Scenario 2: spying phone/video calls



Related Work

Vaidya et al., **Cocaine Noodles** (WOOT 2015)
Carlini et al., **Hidden Voice Commands** (Usenix Security 2016)

The attacking commands are still **audible.**



DolphinAttack

ATTACKED DEVICE : AMAZON ECHO

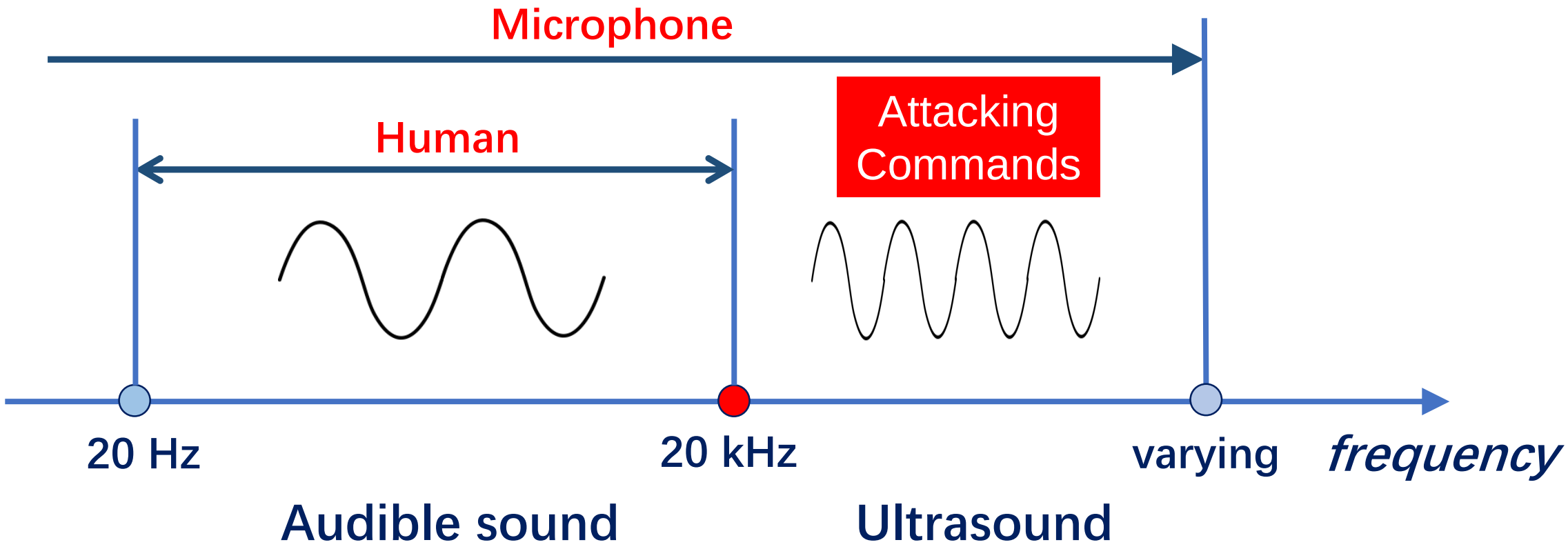
Attack Scenario

- Order stuff
- Make a call
- Read to-do list
- Open the door

Video can be found at usslab.org

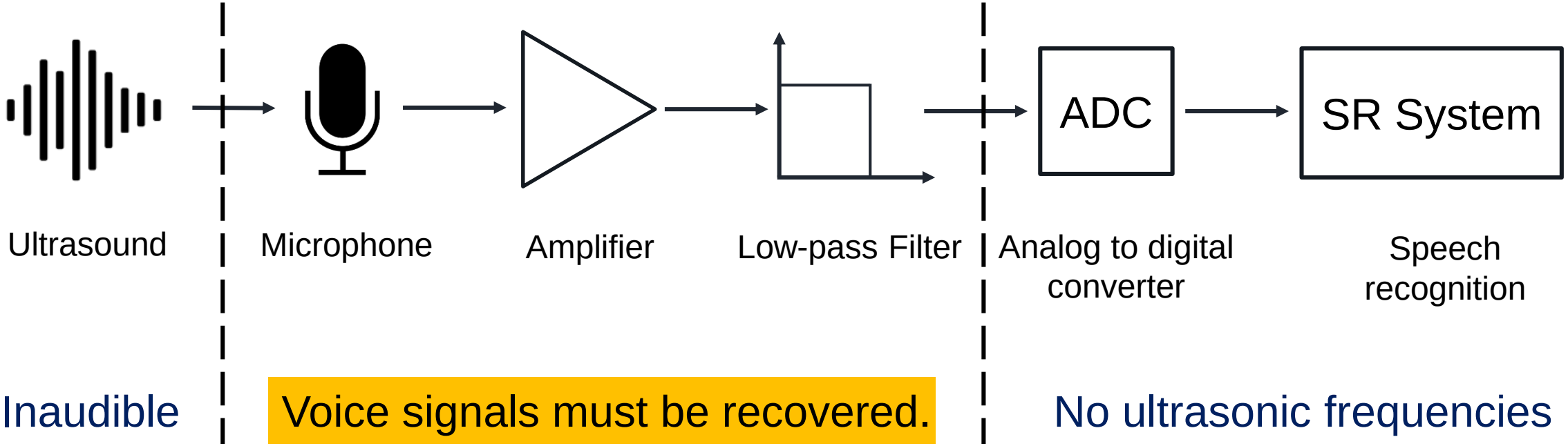
Hearing Range of Human and Microphone

Speech recognition systems only accept signals of **audible sound**.

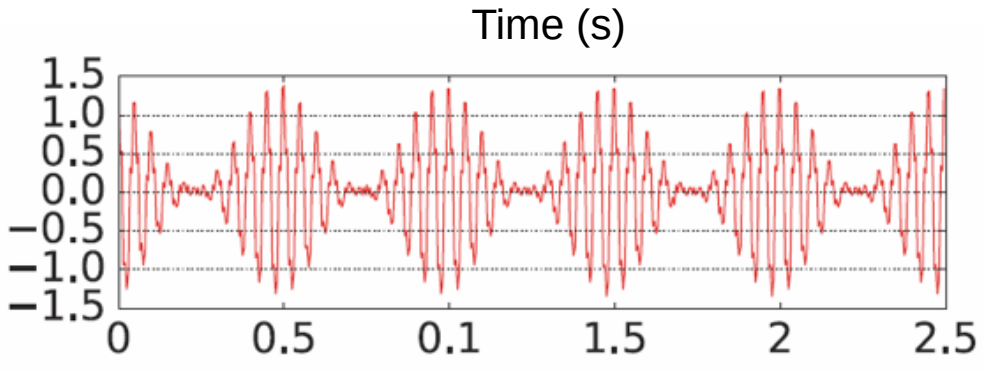


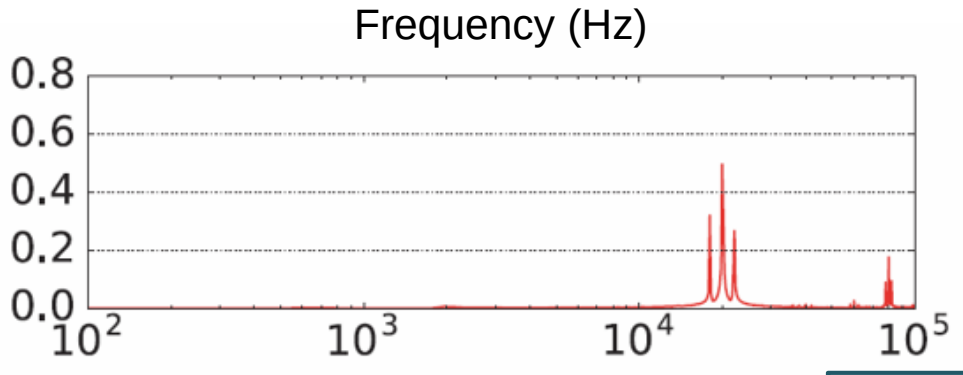
How can voice assistants accept ultrasound?

- The low-pass filter will remove ultrasonic frequencies to avoid aliasing.



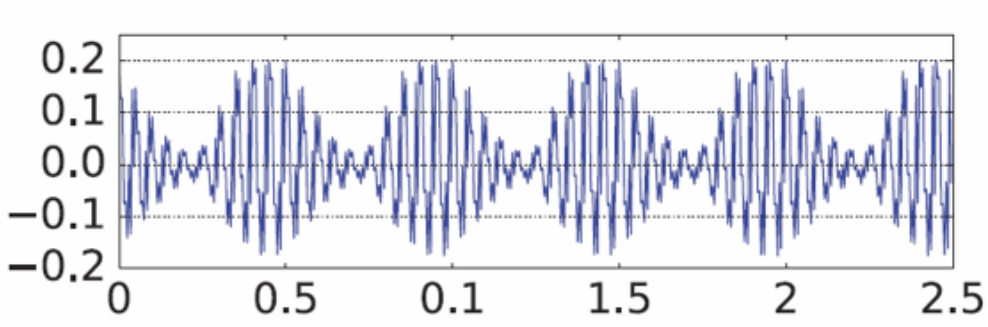
Nonlinearity Effect Validation

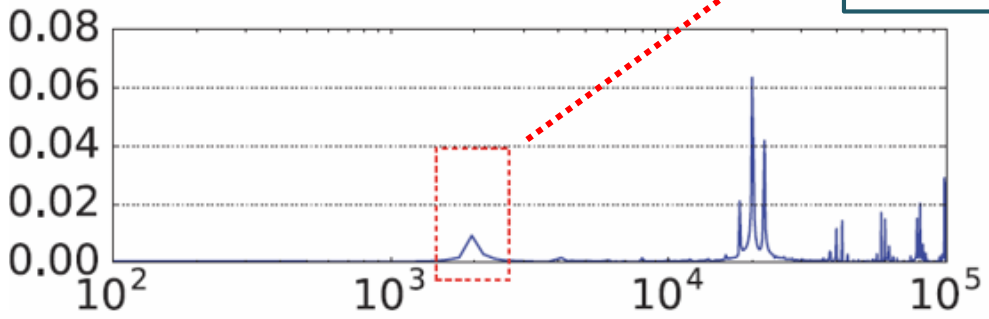




$f_c = 22$ kHz, $f_m = 2$ kHz

Signals of DolphinAttack

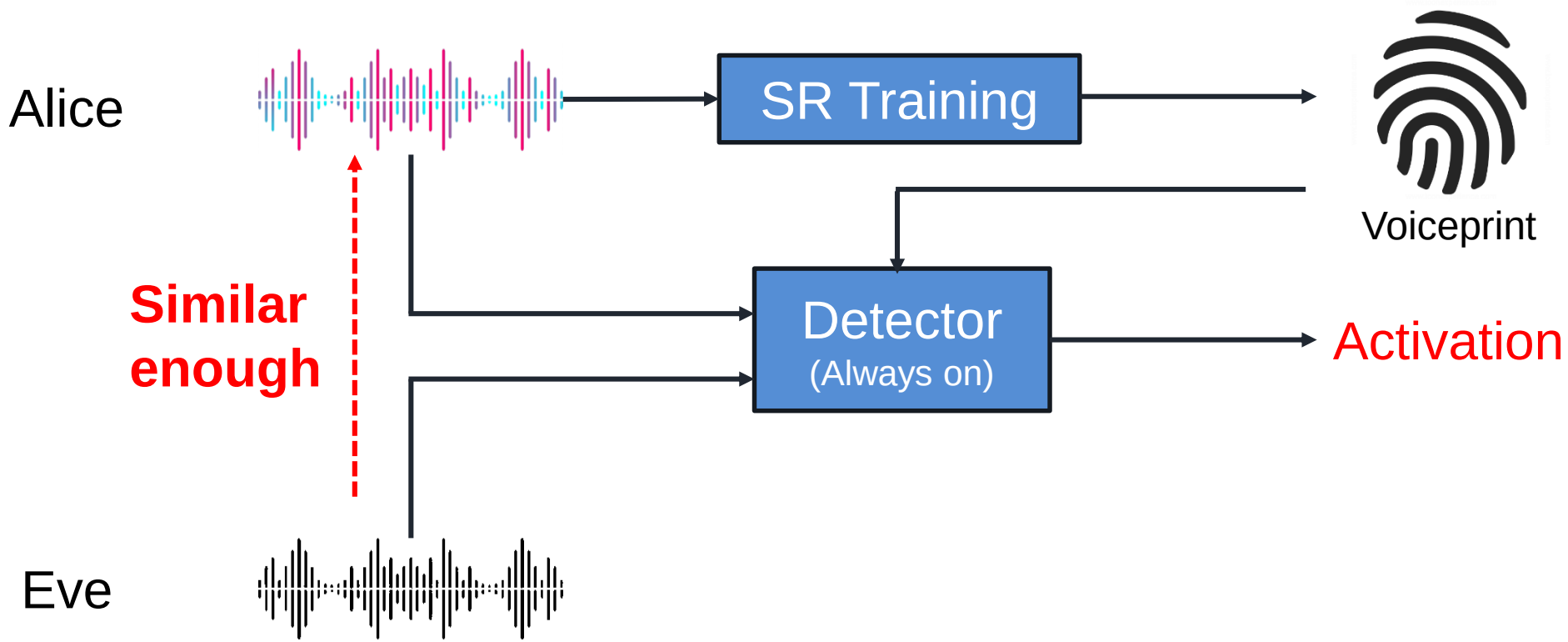




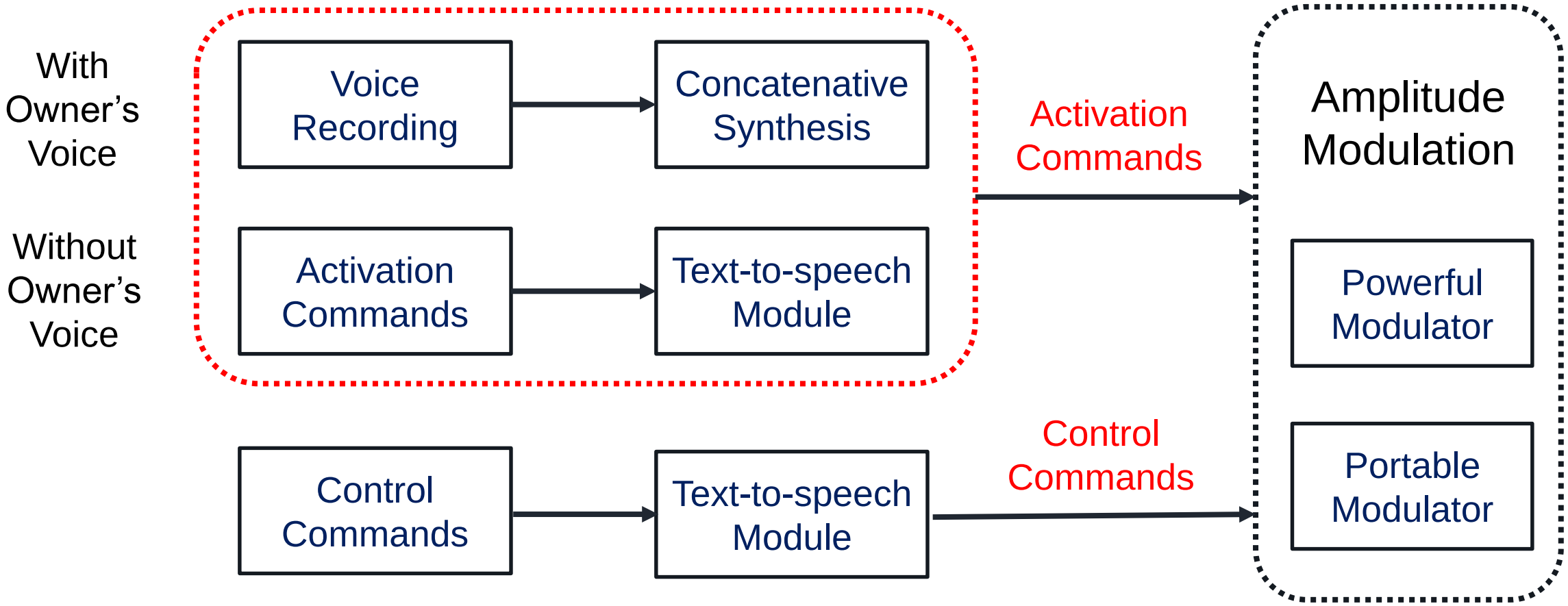
Nonlinearity Effect

Signals received by a MEMS microphone

Speaker Dependent SR – Activation

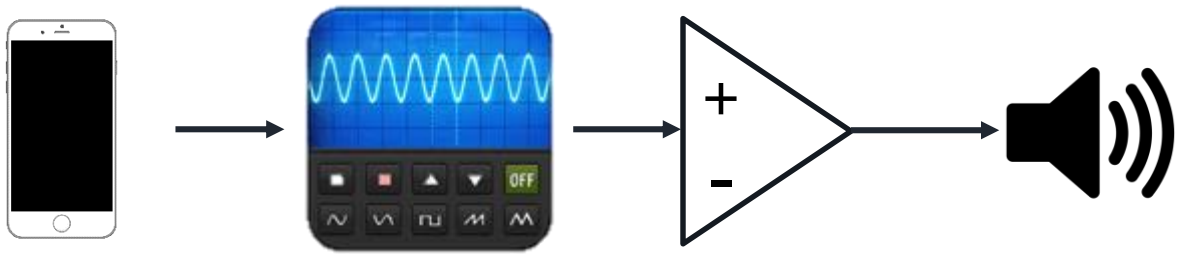


Design of DolphinAttack

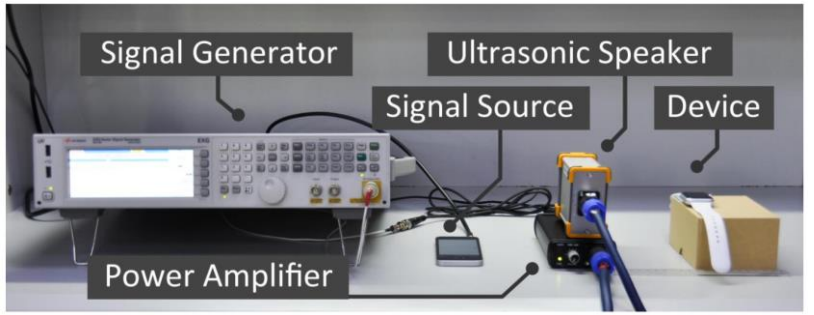


Inaudible Voice Commands Transmitter

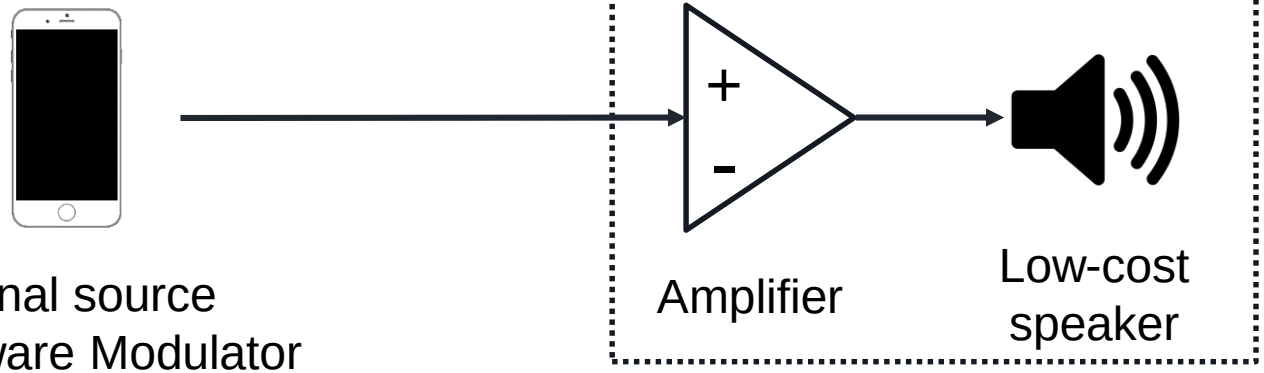
Rich man solution



Signal source | Signal generator (Hardware Modulator) | Amplifier | Professional speaker

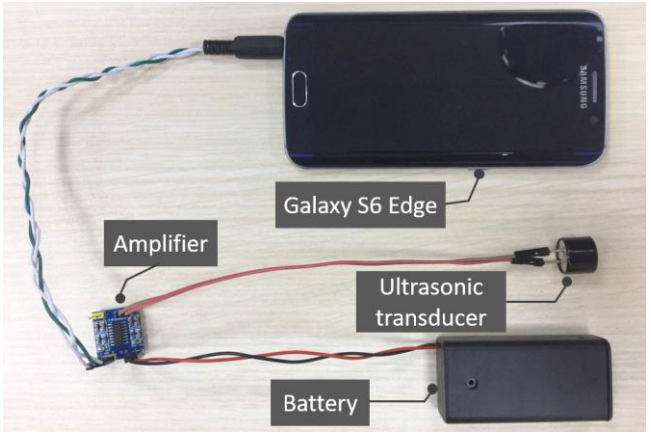


Poor man solution



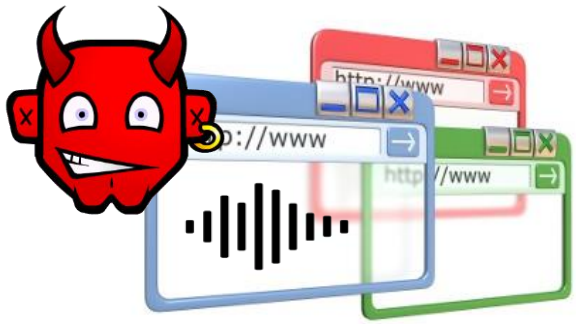
Signal source & Software Modulator

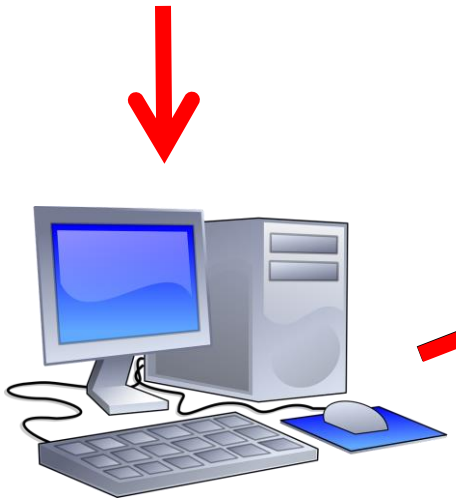
Amplifier | Low-cost speaker



Less than $3

Attack Scenario: Remote Attack





Computer

Commodity Speaker

Smart devices

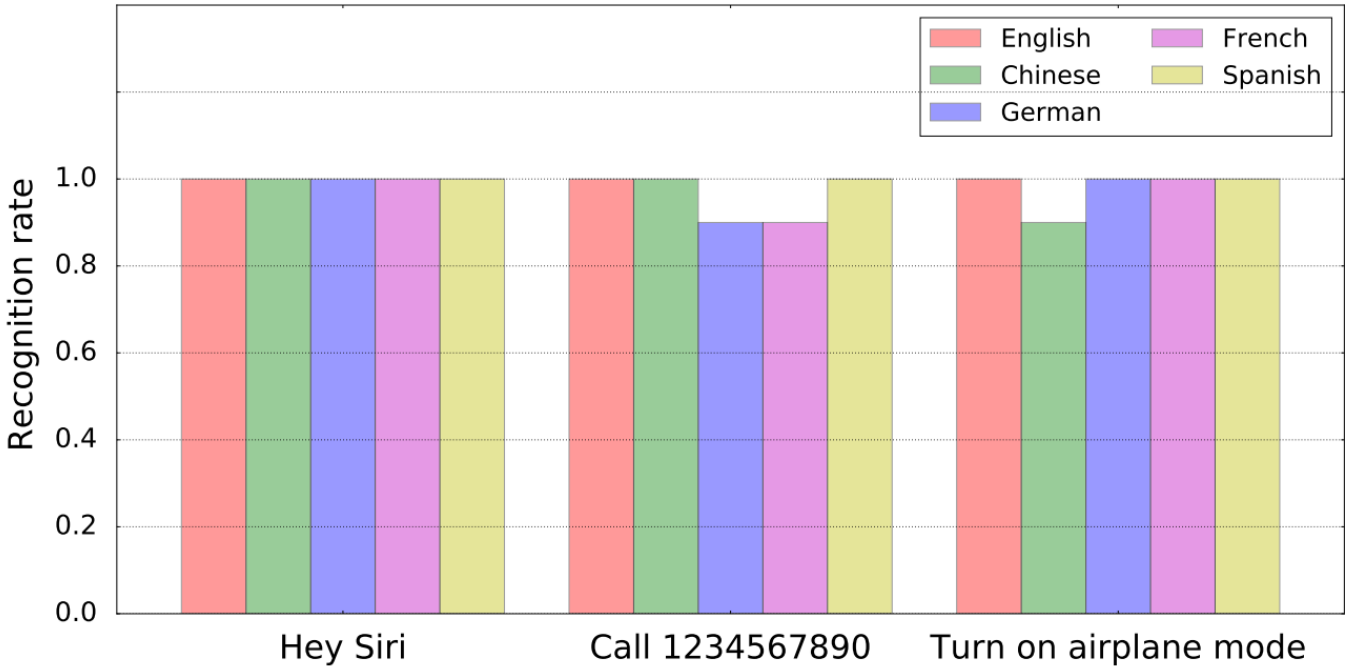
Manuf.	Model	OS/Ver.	SR System	Attacks		Modulation Parameters		Max Dist. (cm)	
				Recog.	Activ.	f_c (kHz) & [Prime f_c] ‡	Depth	Recog.	Activ.
Apple	iPhone 4s	iOS 9.3.5	Siri	√	√	20–42 [27.9]	≥ 9%	175	110
Apple	iPhone 5s	iOS 10.0.2	Siri	√	√	24.1 26.2 27 29.3 [24.1]	100%	7.5	10
Apple	iPhone SE	iOS 10.3.1	Siri	√	√	22–28 33 [22.6]	≥ 47%	30	25
			Chrome	√	N/A	22–26 28 [22.6]	≥ 37%	16	N/A
Apple	iPhone SE †	iOS 10.3.2	Siri	√	√	21–29 31 33 [22.4]	≥ 43%	21	24
Apple	iPhone 6s *	iOS 10.2.1	Siri	√	√	26 [26]	100%	4	12
Apple	iPhone 6 Plus *	iOS 10.3.1	Siri	×	√	— [24]	—	—	2
Apple	iPhone 7 Plus *	iOS 10.3.1	Siri	√	√	21 24-29 [25.3]	≥ 50%	18	12
Apple	watch	watchOS 3.1	Siri	√	√	20–37 [22.3]	≥ 5%	111	164
Apple	iPad mini 4	iOS 10.2.1	Siri	√	√	22–40 [28.8]	≥ 25%	91.6	50.5
Apple	MacBook	macOS Sierra	Siri	√	N/A	20-22 24-25 27-37 39 [22.8]	≥ 76%	31	N/A
LG	Nexus 5X	Android 7.1.1	Google Now	√	√	30.7 [30.7]	100%	6	11
Asus	Nexus 7	Android 6.0.1	Google Now	√	√	24–39 [24.1]	≥ 5%	88	87
Samsung	Galaxy S6 edge	Android 6.0.1	S Voice	√	√	20–38 [28.4]	≥ 17%	36.1	56.2
Huawei	Honor 7	Android 6.0	HiVoice	√	√	29–37 [29.5]	≥ 17%	13	14
Lenovo	ThinkPad T440p	Windows 10	Cortana	√	√	23.4–29 [23.6]	≥ 35%	58	8
Amazon	Echo *	5589	Alexa	√	√	20-21 23-31 33-34 [24]	≥ 20%	165	165
Audi	Q3	N/A	N/A	√	N/A	21–23 [22]	100%	10	N/A

‡ Prime f_c is the carrier wave frequency that exhibits highest baseband amplitude after demodulation.
† Another iPhone SE with identical technical spec.
* Experimented with the front/top microphones on devices.
— No result

Evaluation: Impact of Languages

DolphinAttack is effective for various languages and voice commands.

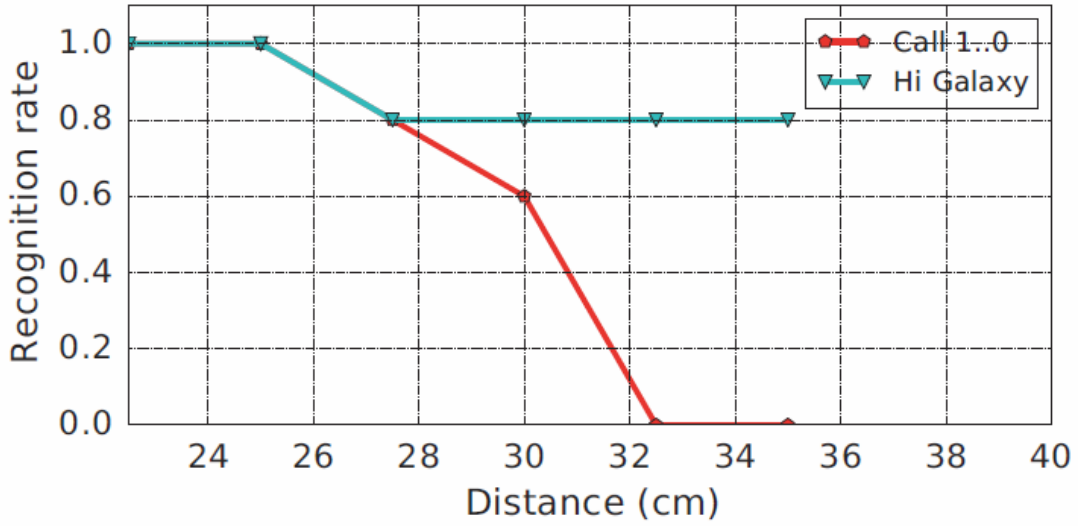
- English
- Chinese
- French
- German
- Spanish



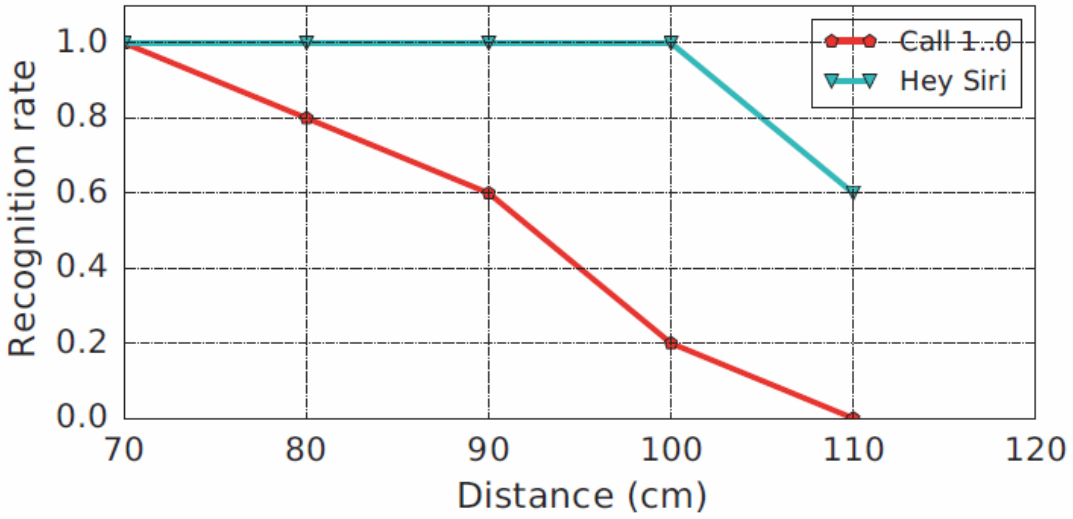
The recognition rates of voice commands in five languages

Evaluation: Impact of Attack Distance

The attack distance has fundamental impact on the effectiveness of DolphinAttack and is device dependent.



(a) The recognition rates of the Galaxy S6 Edge

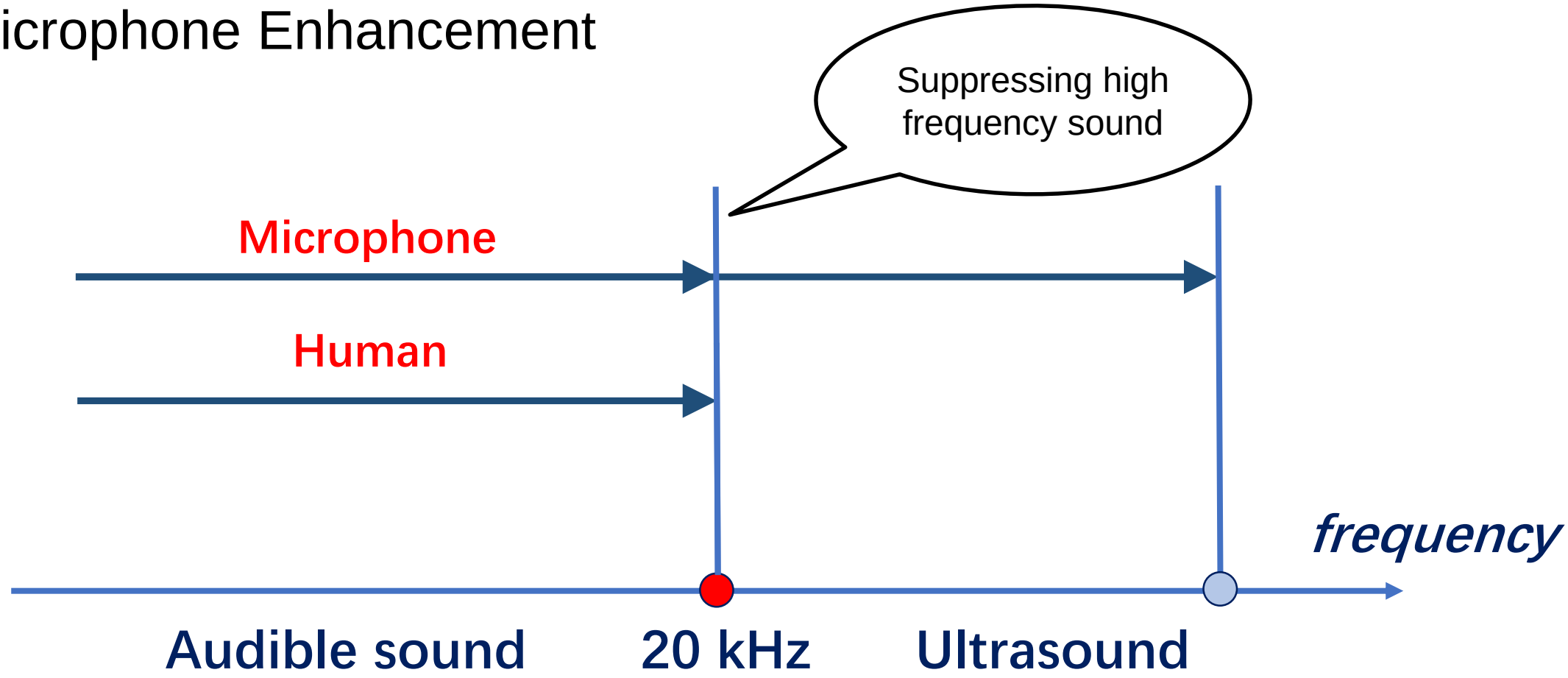


(b) The recognition rates of the Apple watch

The impact of attack distances on the recognition rates for S6 Edge and Apple watch.

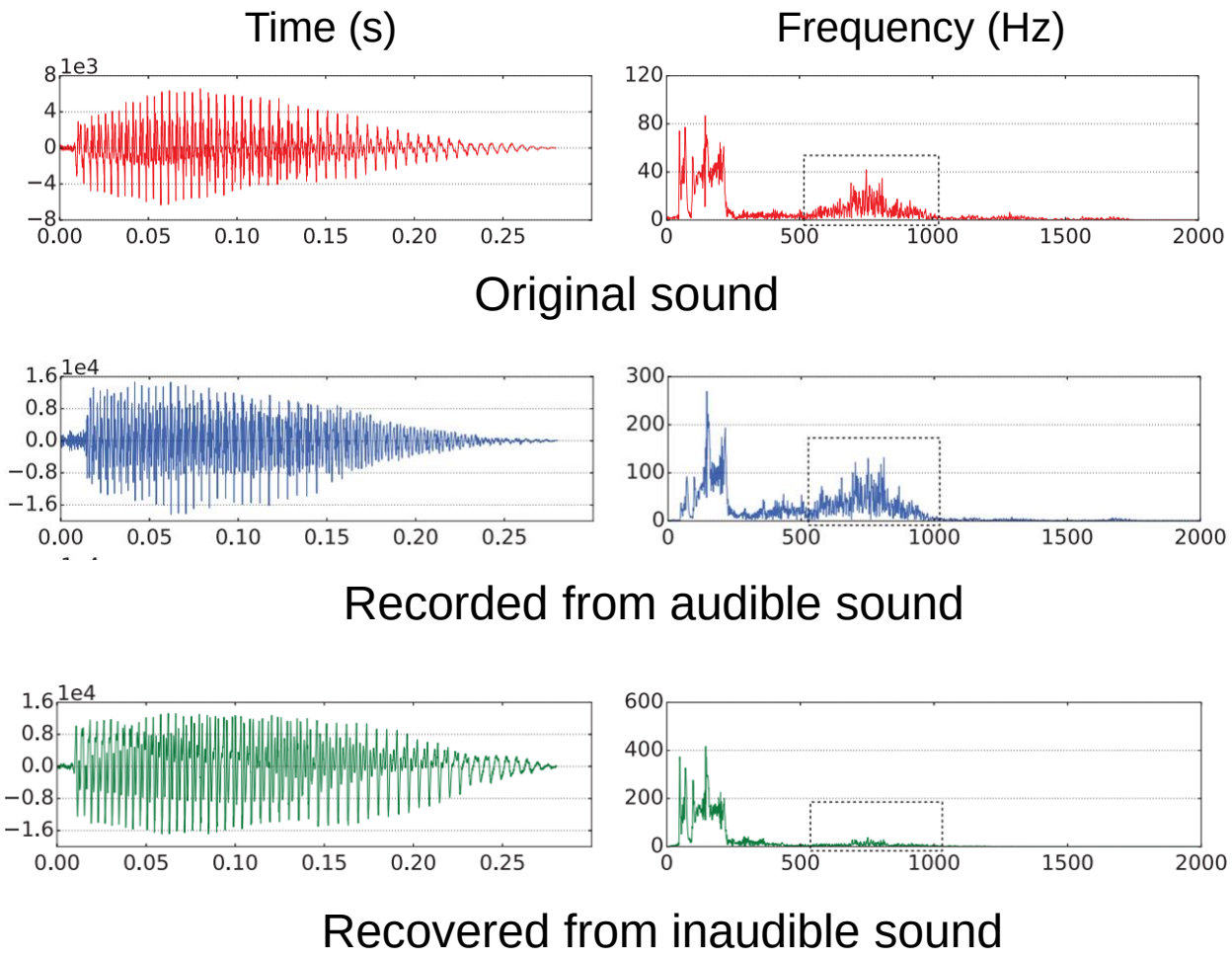
Hardware-Based Defense

- Microphone Enhancement



Software-Based Defense

- Modulated voice commands are distinctive from genuine ones.
- Supported vector machine (SVM) as the classifier to detect the malicious command from the normal command.
- Result: 100% true positive rate (7/7) and 100% true negative rate (7/7).



Summary

- Voice assistant has become an increasingly popular human-computer interaction mechanism, but they are vulnerable to attacks.
- DolphinAttack is a totally inaudible attack from a new perspective, could attack Siri, Alexa, Google Now, Cortana, Samsung S Voice, Huawei Hi Voice.
- To avoid the abuse of DolphinAttack in reality, we propose two defense solutions from the aspects of both hardware and software.

Questions

DolphinAttack Homepage: http://dolphinattack.com/

USS Lab Homepage: http://usslab.org/