DOLPHIN ATTACK

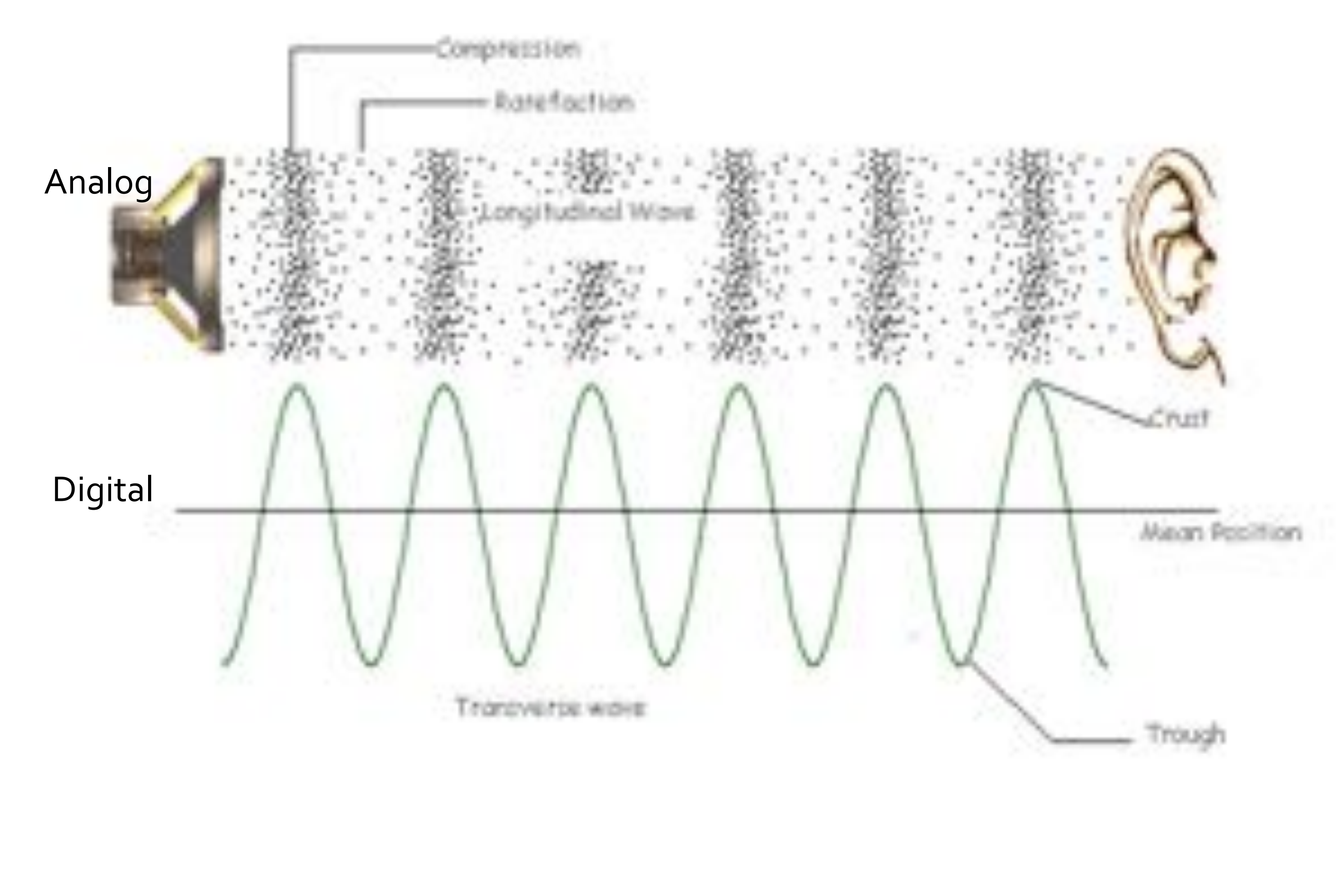
GUOMING ZHANG, CHEN YAN

PRESENTED BY JACOB BEDNARD
WAYNE STATE UNIVERSITY CSC6991

Overview

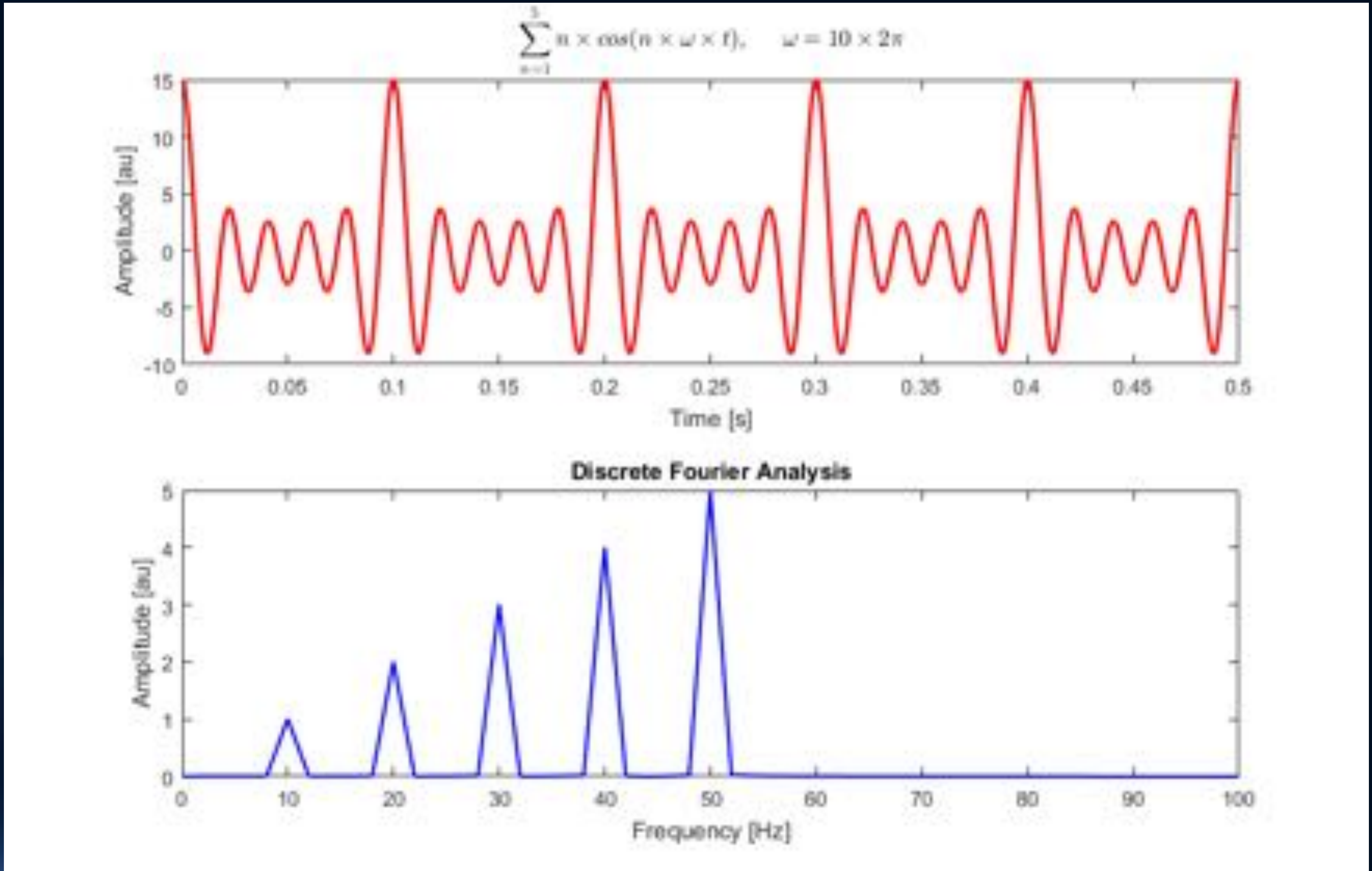
- Soundwaves and Digital Signal Processing (DSP)
- Attack Methodology
- Defense Mechanisms
- Demonstration Videos

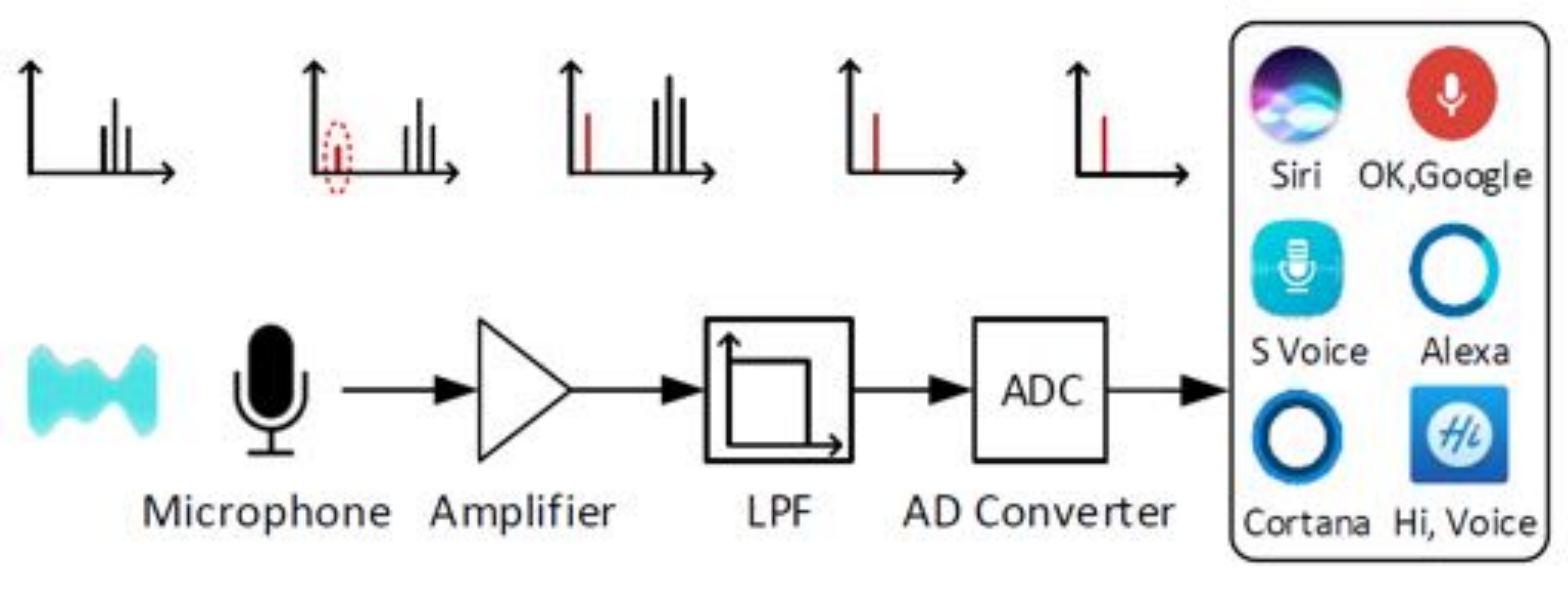
Soundwaves and Digital Signal Processing (DSP)

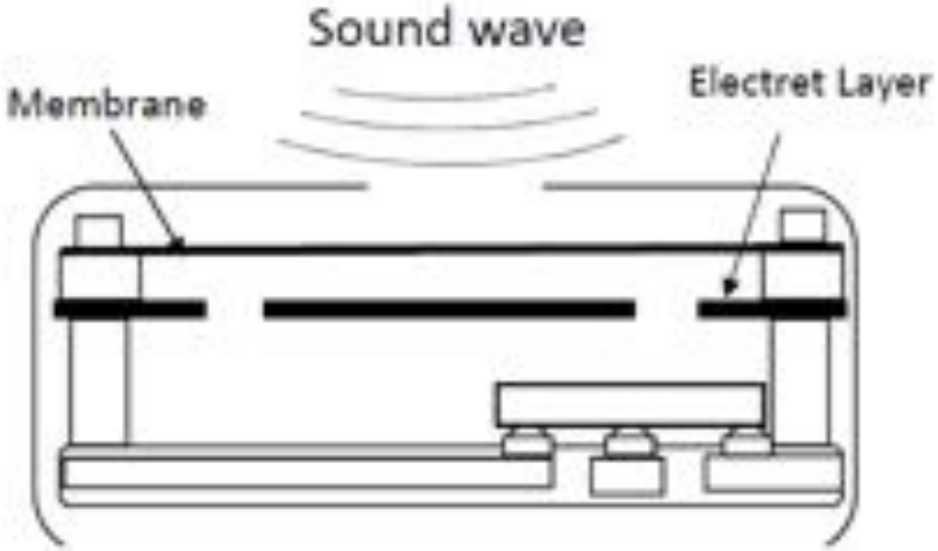


Analog

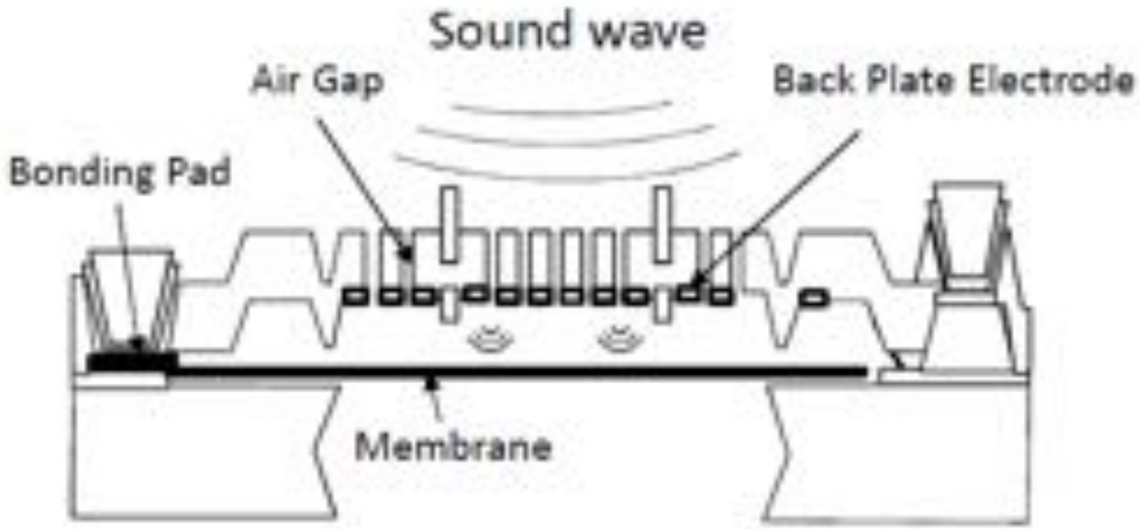
Digital


$\sum_{n=1}^{5} n \times \cos(n \times \omega \times f), \quad \omega = 10 \times 2\pi$
Amplitude [au]
15
10
5
0
-5
-10
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
Time [s]
Discrete Fourier Analysis
Amplitude [au]
5
4
3
2
1
0
0
10
20
30
40
50
60
70
80
90
100
Frequency [Hz]

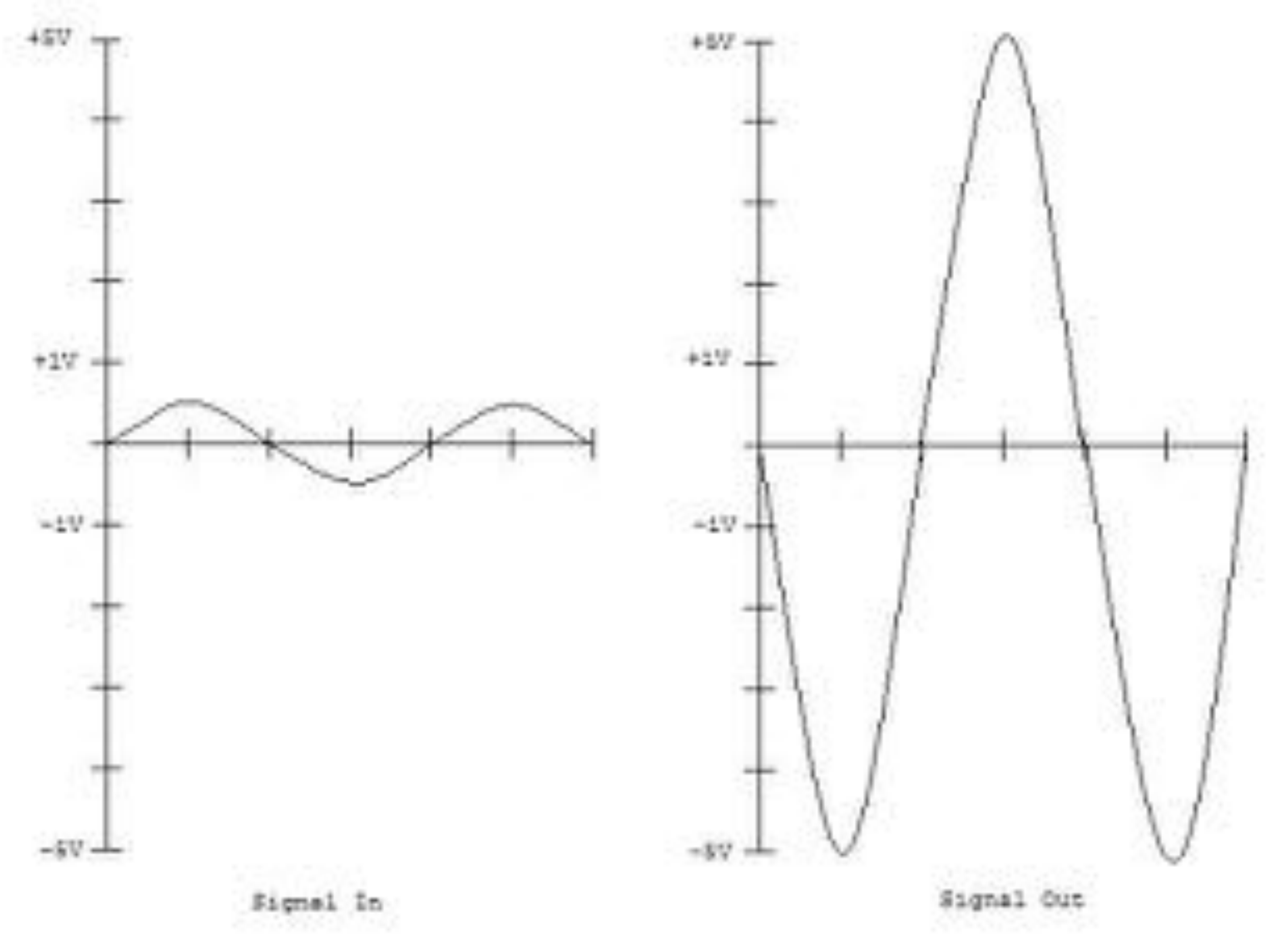

Microphone
Amplifier
LPF
ADC
AD Converter
Siri
OK,Google
S Voice
Alexa
Cortana
Hi, Voice



(a) Structure of ECM



(b) Structure of MEMS Microphone


+5V
+1V
-1V
-5V
Signal In
+5V
+1V
-1V
-5V
Signal Out

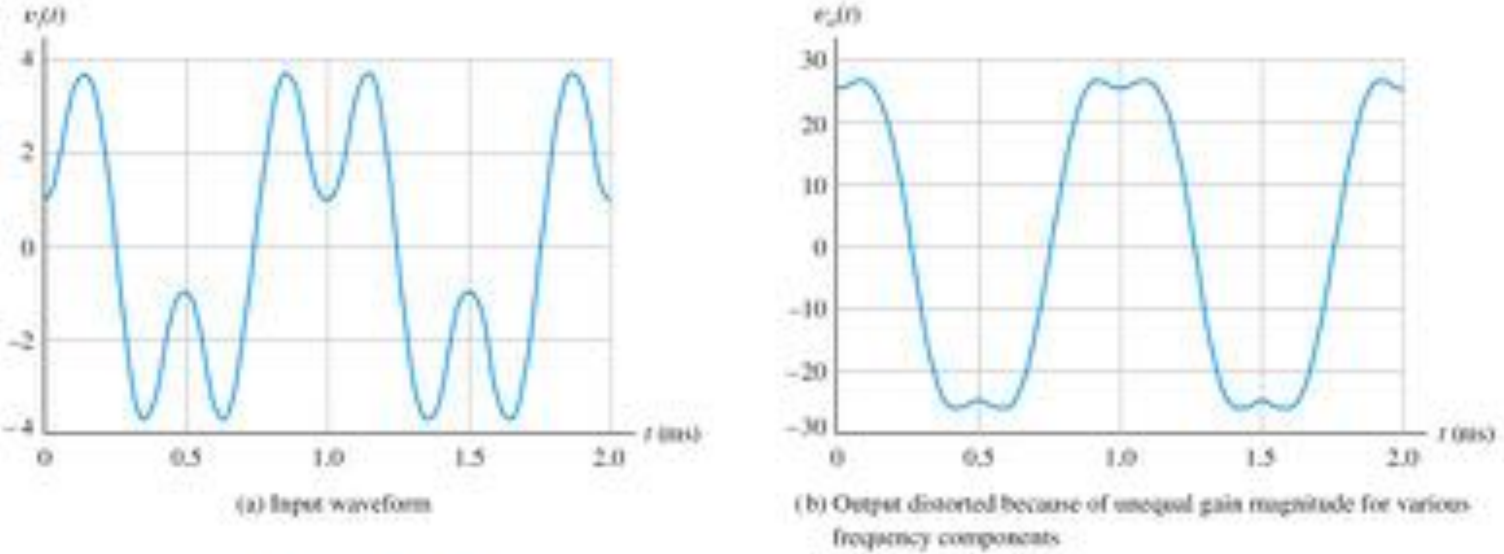
11. Amplifiers – Linear Waveform Distortion

11.8 Linear Waveform Distortion

* *Distortion* may occur even though the amplifier is *linear* (i.e., obeys superposition principle).

Amplitude Distortion

If a signal contains components of various frequencies, the output waveform may be distorted due to the frequency response of the amplifier gain.

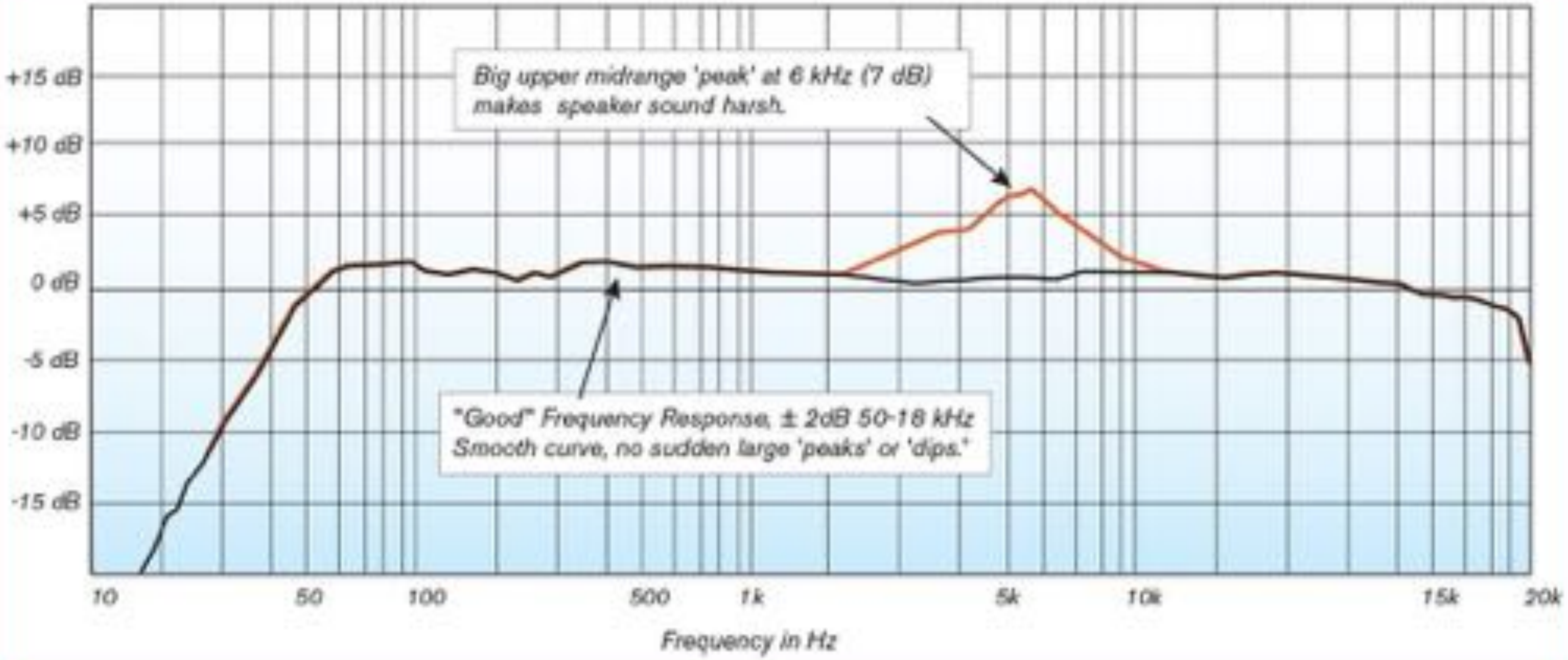


(a) Input waveform

(b) Output distorted because of unequal gain magnitude for various frequency components

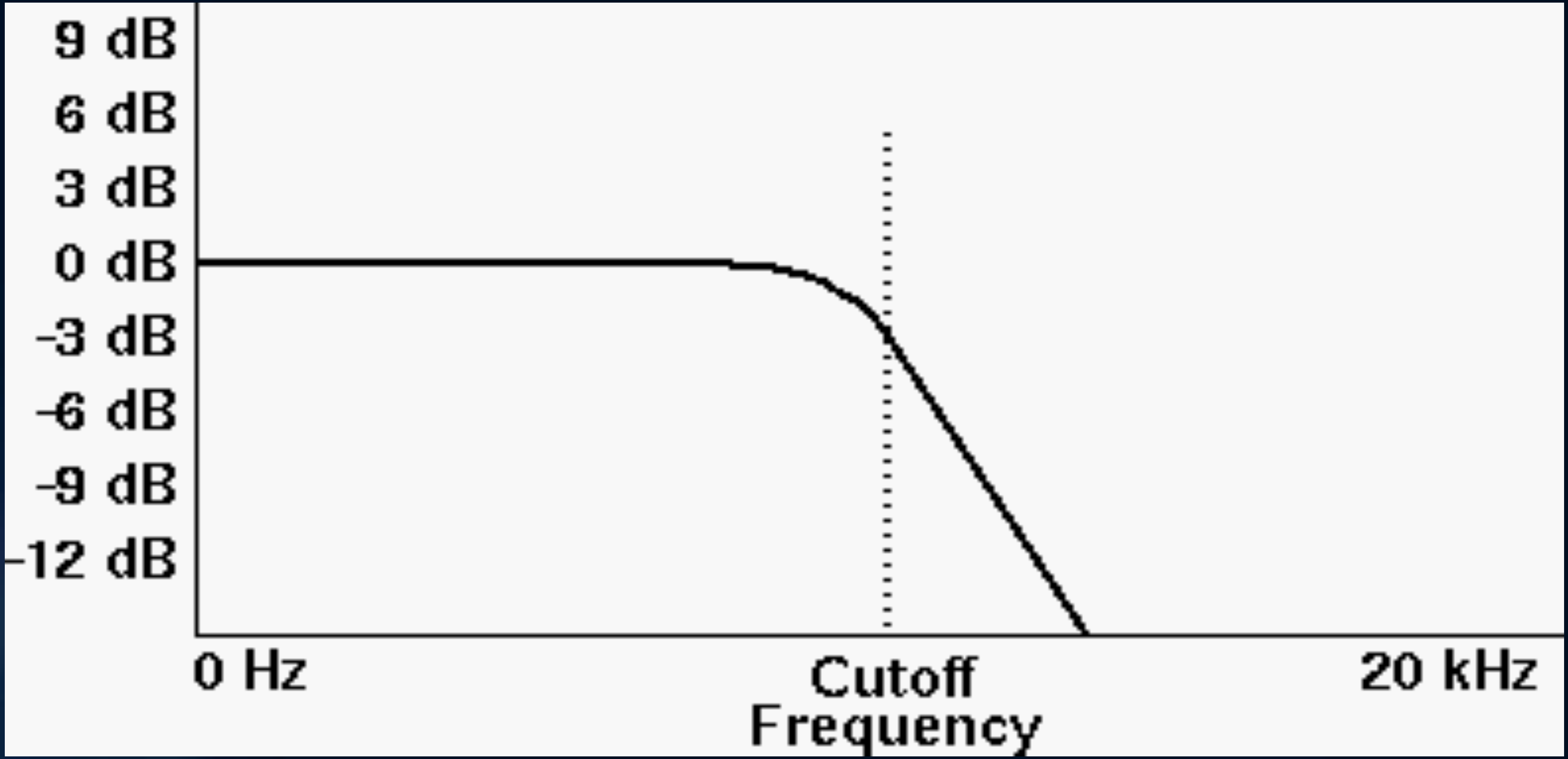
Figure 11.26 Linear amplitude distortion. See Example 11.9.

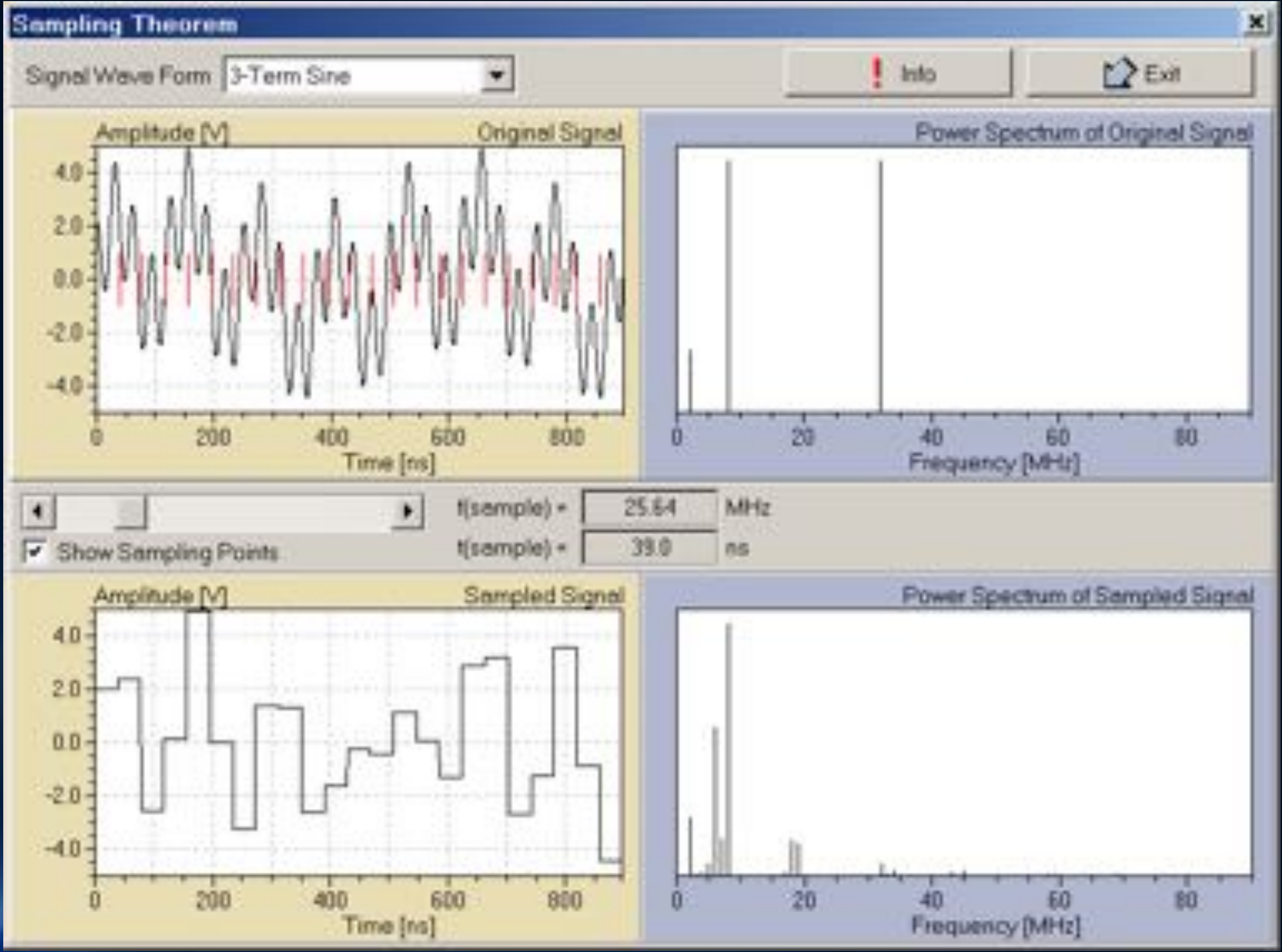
Figure 2 Speaker frequency responses: Good and bad



Big upper midrange 'peak' at 6 kHz (7 dB) makes speaker sound harsh.

"Good" Frequency Response, ± 2dB 50-18 kHz Smooth curve, no sudden large 'peaks' or 'dips.'


9 dB
6 dB
3 dB
0 dB
-3 dB
-6 dB
-9 dB
-12 dB
0 Hz
Cutoff
Frequency
20 kHz



Signal Wave Form: 3-Term Sine

f(sample) = 25.64 MHz

t(sample) = 39.0 ns

☑ Show Sampling Points

Attack Methodology

Dolphin Attack Major Contributions

- Attackers can use inaudible sounds (>20kHz) to activate and control voice controllable systems such as cell phones, home entertainment systems, vehicles, etc.
- The attack is sneaky in nature. Device owners may not be aware that they are being attacked due to the remote distance and inaudible sounds that the attack utilizes.

Dolphin Attack…?

What does this have to do with Dolphins?

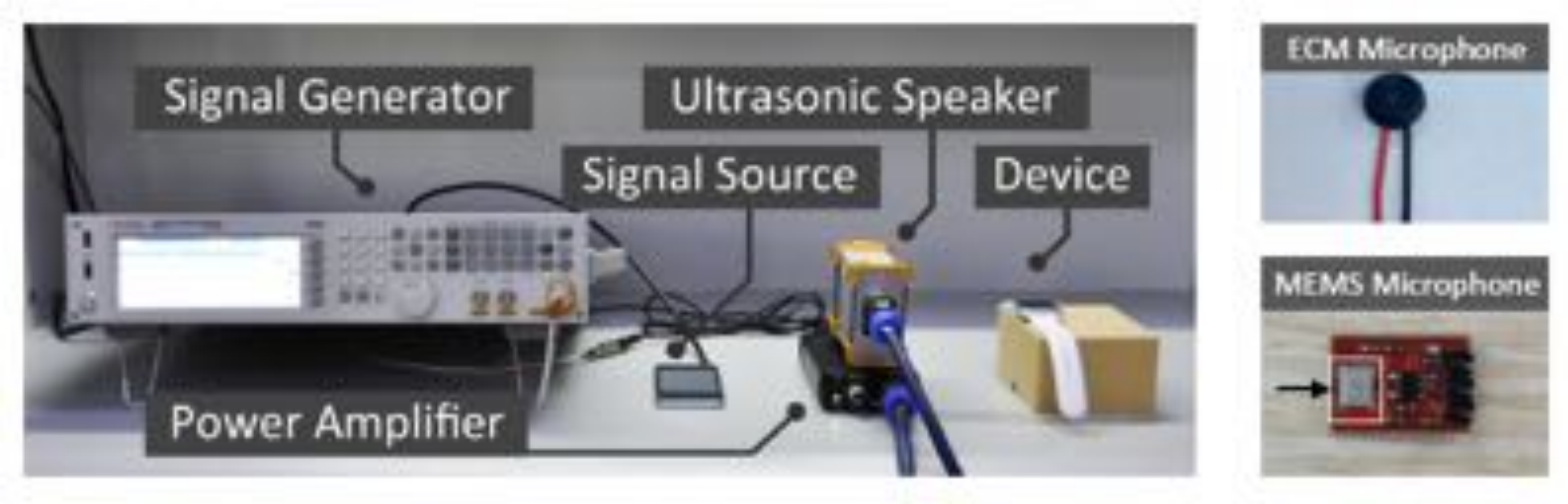
Motivation

- Eaves Dropping (Leak Personal Data, Authentication By-Pass, etc.)
- “Sneaky” Attack
- Defeat the “Air-Gap”
- Drive-By Attack

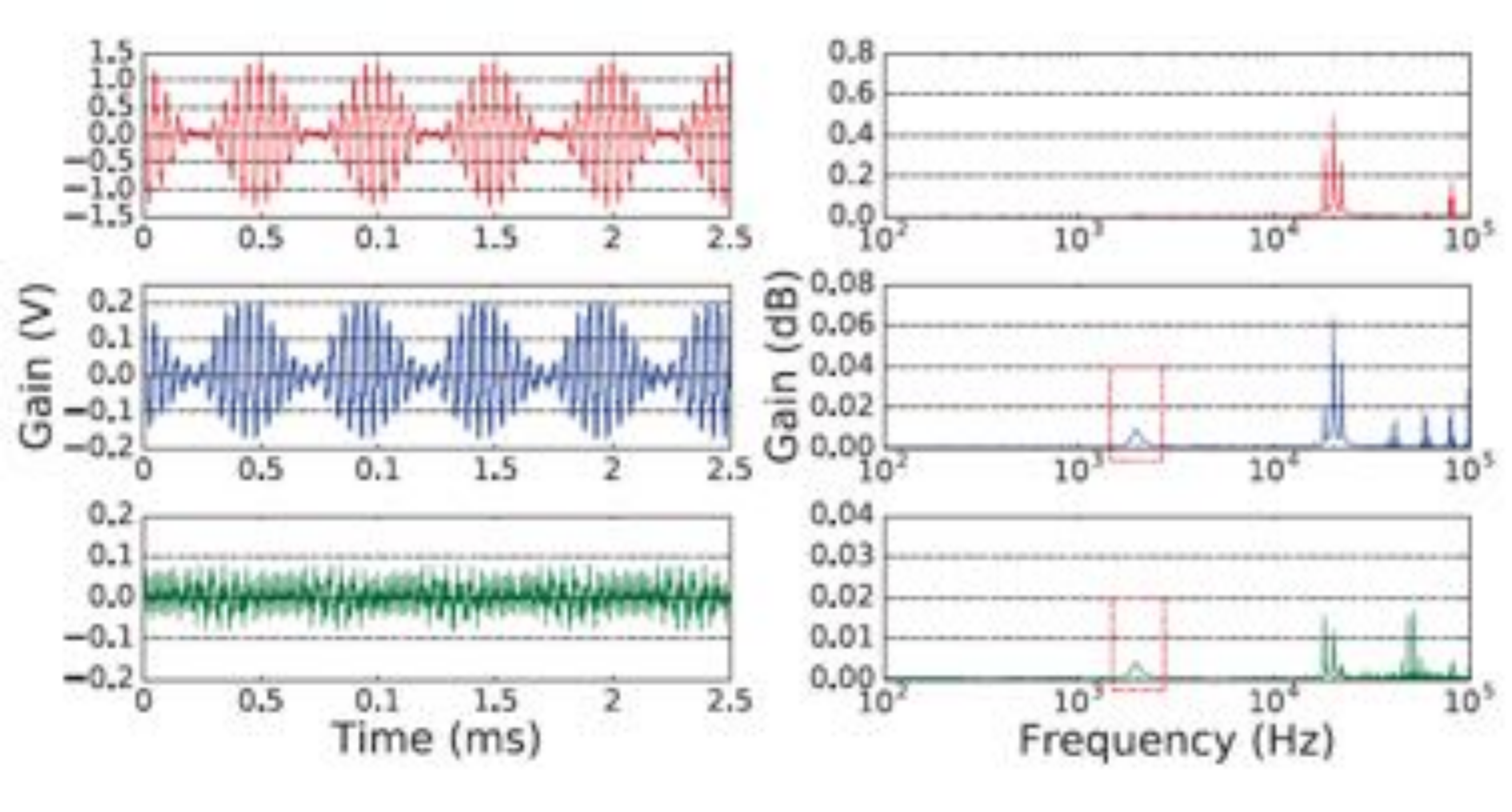
Threat Model

- No Target Device Access
- No Owner Interaction
- Inaudible
- Attacking Equipment

Feasibility Analysis (Test Setup)



Feasibility Analysis (Exploiting Linear Amps)



Feasibility Analysis (Speakers & Microphones)

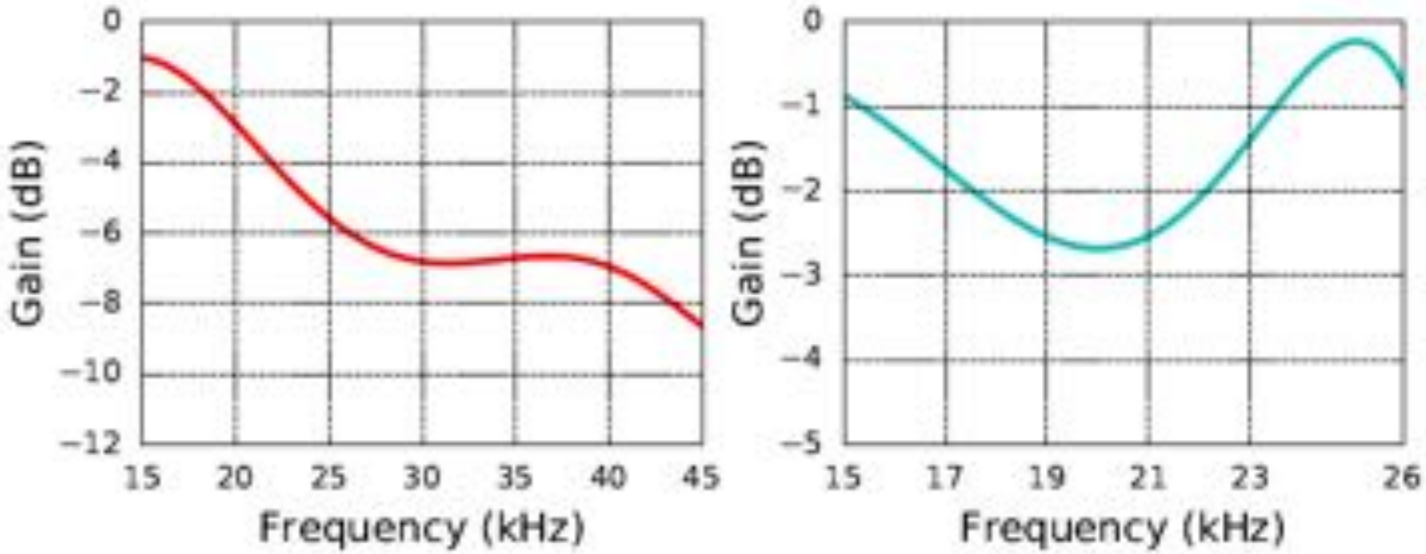
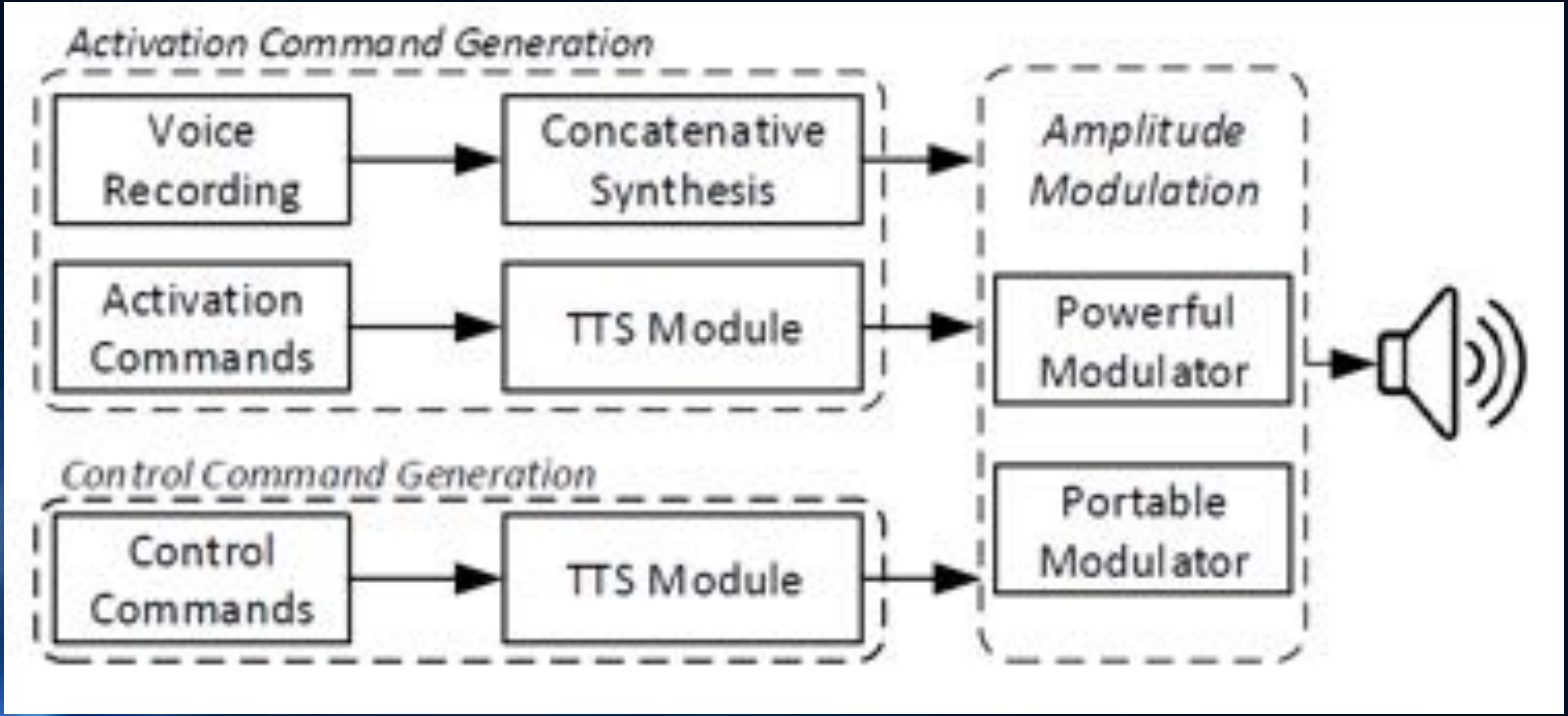
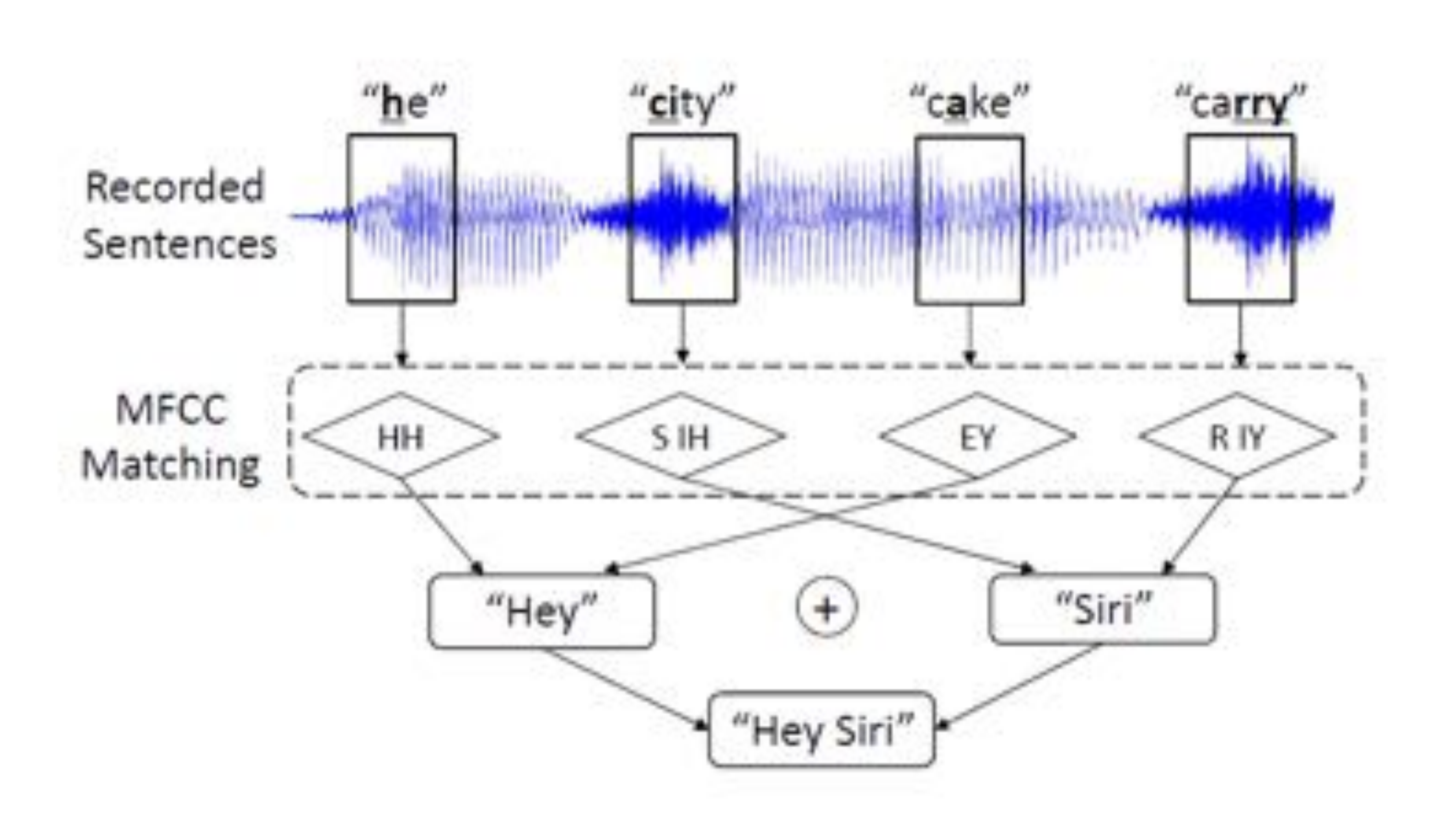


Figure 10: The frequency responses of the ADMP401 MEMS microphone (left) and the Samsung Galaxy S6 Edge speaker (right).

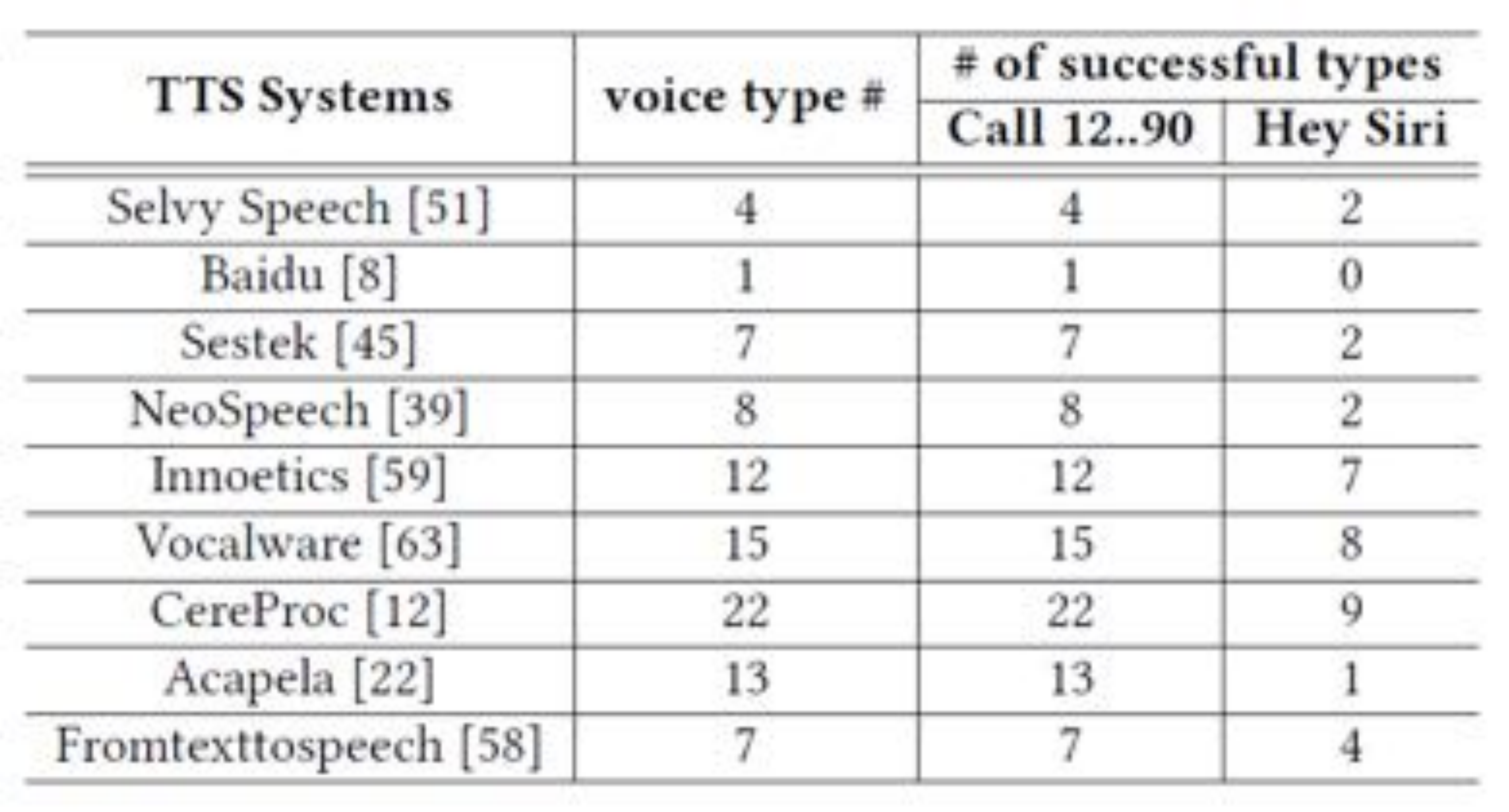
Attack Design



Human Voice Samples

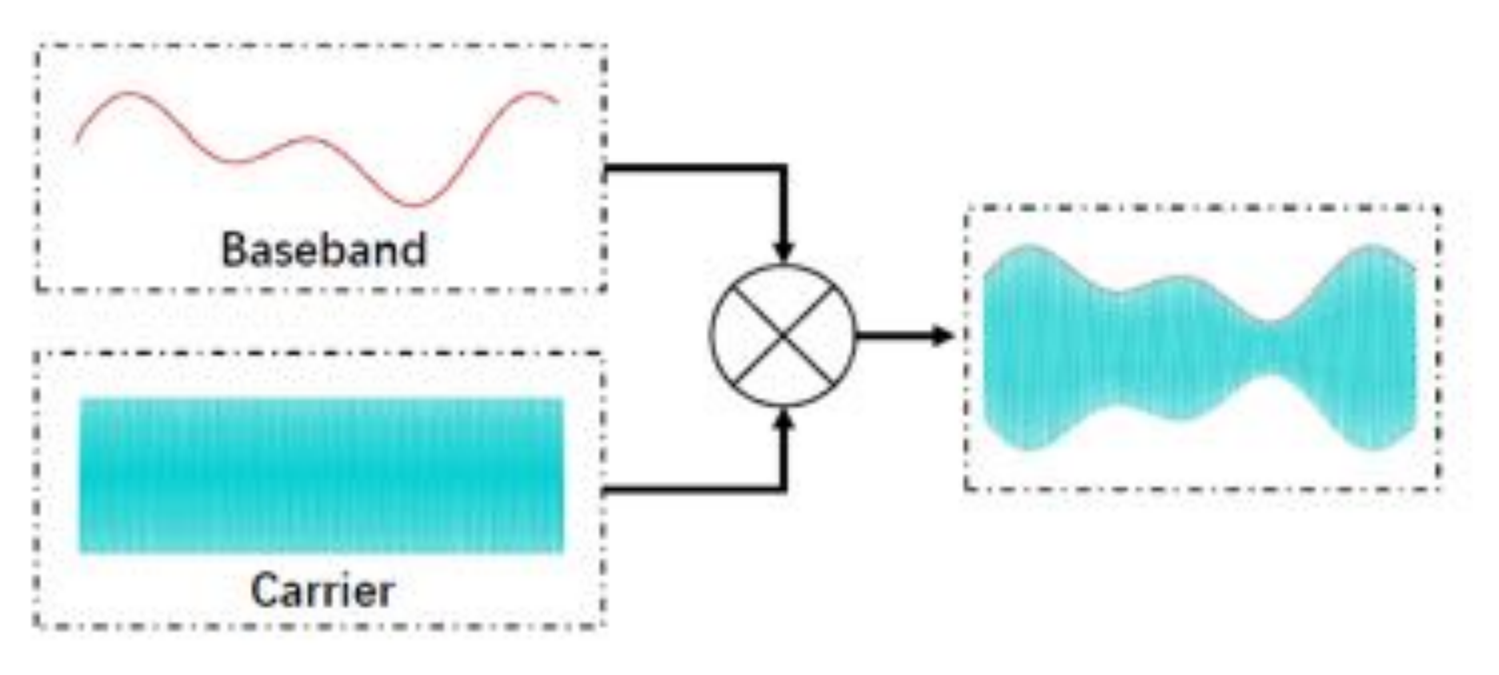


Text-to-Speech Samples

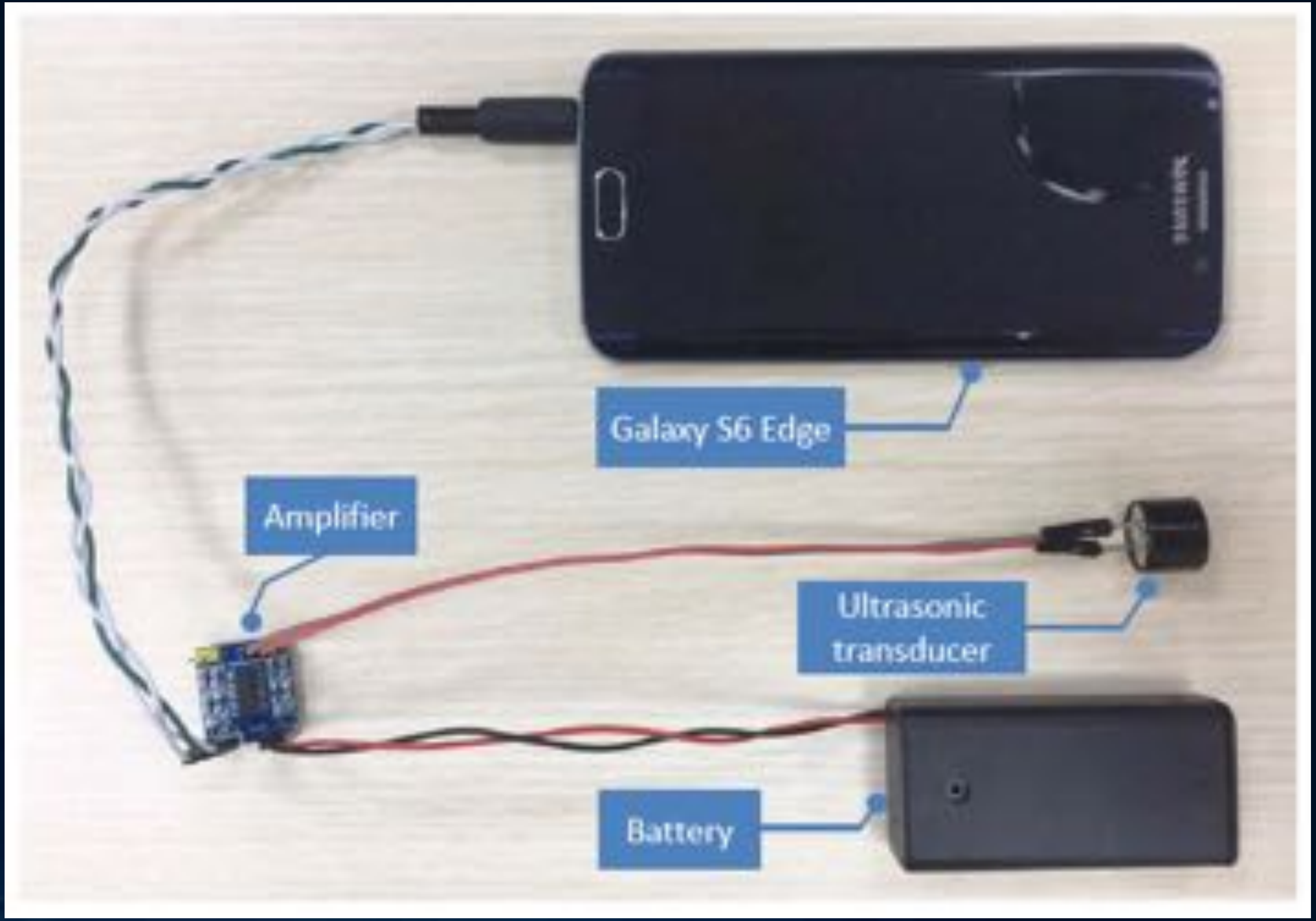


TTS Systems	voice type #	# of successful types	
		Call 12..90	Hey Siri
Selvy Speech [51]	4	4	2
Baidu [8]	1	1	0
Sestek [45]	7	7	2
NeoSpeech [39]	8	8	2
Innoetics [59]	12	12	7
Vocalware [63]	15	15	8
CereProc [12]	22	22	9
Acapela [22]	13	13	1
Fromtexttospeech [58]	7	7	4

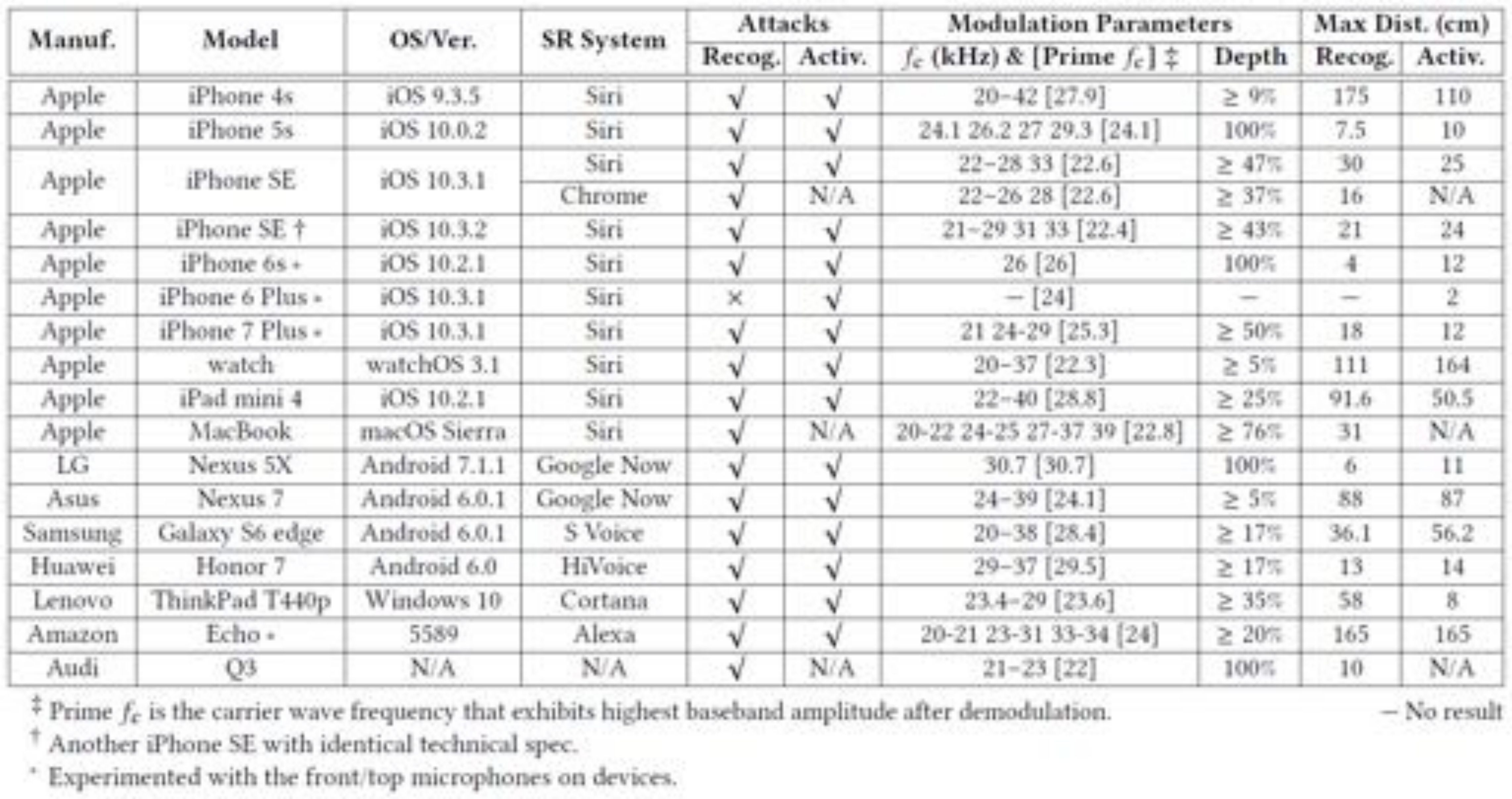
Command Modulation (Conversion to Inaudible)



Low-Cost Attack Implementation



Test Results



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. – No result

† Another iPhone SE with identical technical spec.

* Experimented with the front/top microphones on devices.

Test Results

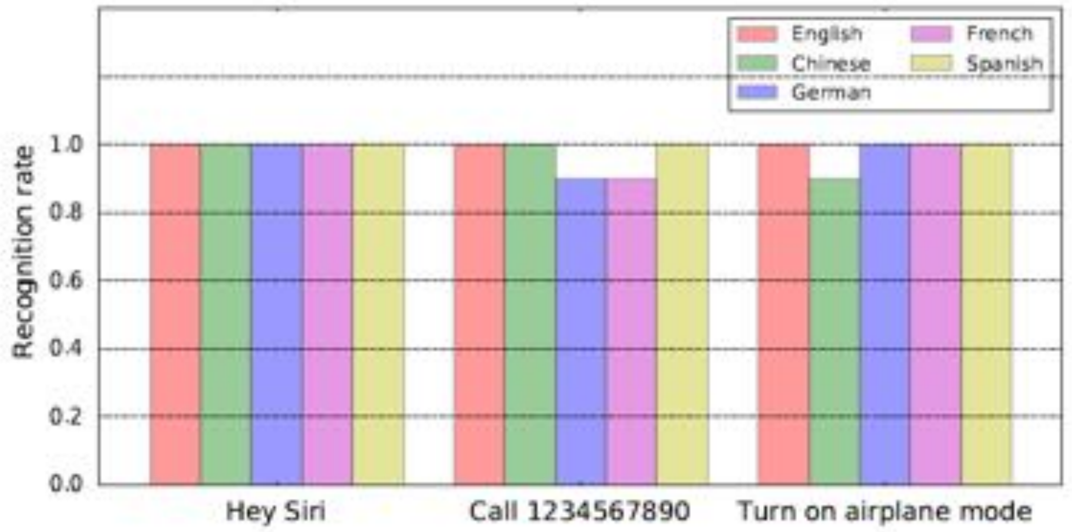
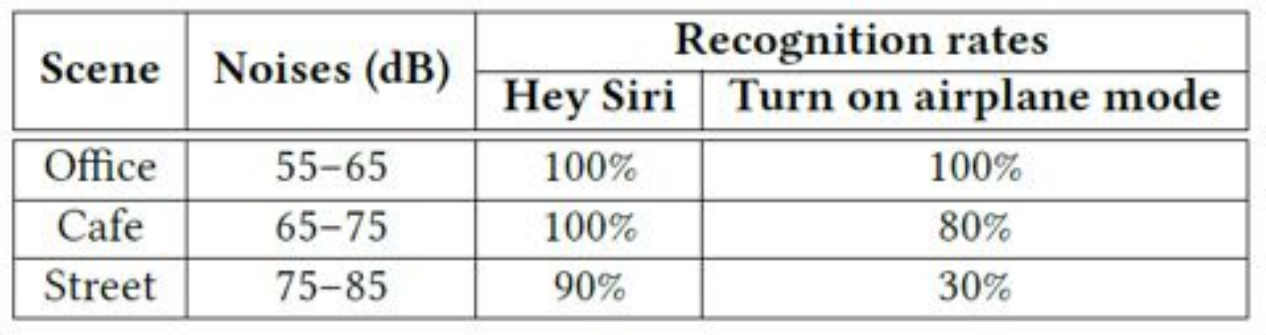


Figure 14: The recognition rates of voice commands in five languages.



Scene	Noises (dB)	Recognition rates	
		Hey Siri	Turn on airplane mode
Office	55–65	100%	100%
Cafe	65–75	100%	80%
Street	75–85	90%	30%

Defense Mechanisms

Hardware-Based Defenses

- Microphone Enhancement
- Inaudible Voice Command Cancellation

Software-Based Defenses

- Supported Vector Machines
- (Machine learning)
- 100% Success Rate w/ 24 voice samples

Demonstration Videos

http://usslab.org/projects/dolphinAttack.html

Summary

- Soundwaves and Digital Signal Processing (DSP)
- Attack Methodology
- Defense Mechanisms
- Demonstration Videos

Questions / Comments?