Inaudible Voice Commands: The Long-Range Attack and Defense

Nirupam Roy, Sheng Sen, Haitham Hassanieh, Romit Roy Choudhury

Presented By: **Shivam Bharuka**

When is your bedtime?

Tell me a joke

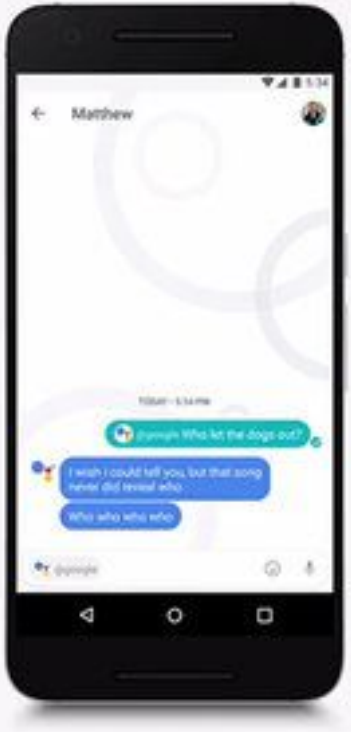
What does the fox say?

Send me a poem

Do I have any photos of cats?

Cheap breakfast options?

What time is it in Beijing?



@google Who let the dogs out?

I wish I could tell you but that song never did reveal who

Who who who who

Where do you live?

Find me cute dog videos

Are you my friend?

Add the Google 10/4 event

Show me the news today

What is the meaning of life?

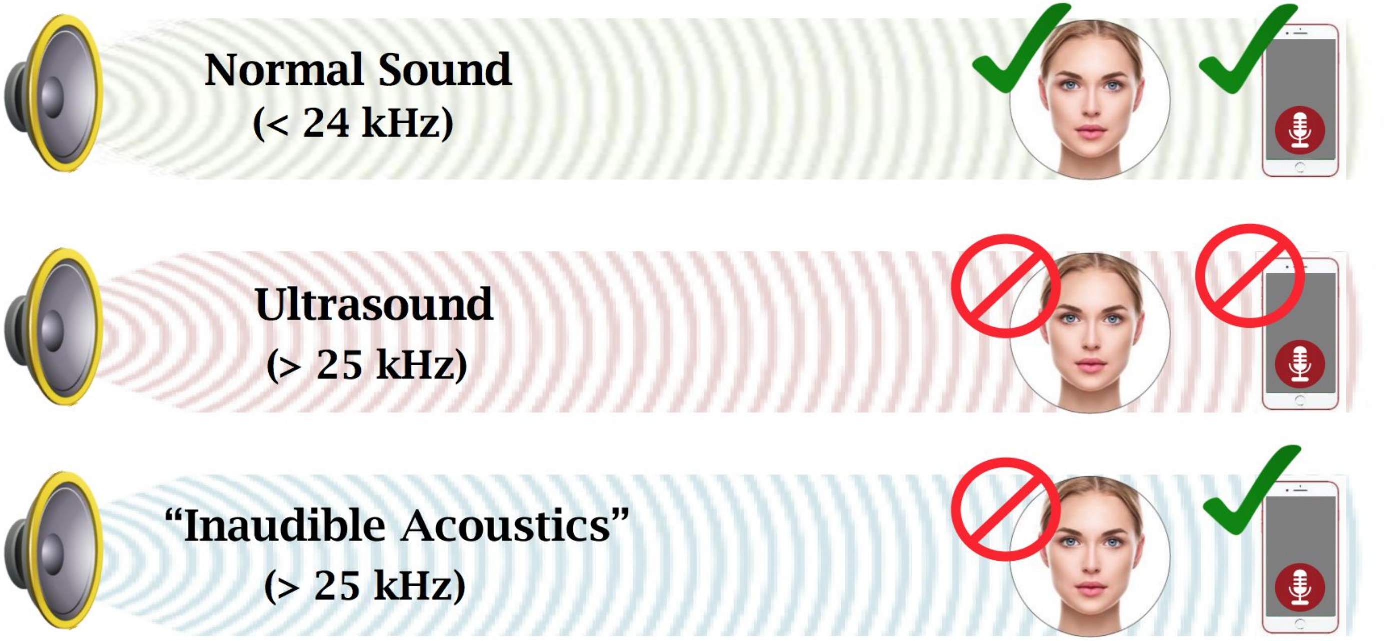
Do you speak morse code?

Who let the dogs out?

Show me high resolution photos of fruit floating threateningly at night

Inaudible Voice Commands

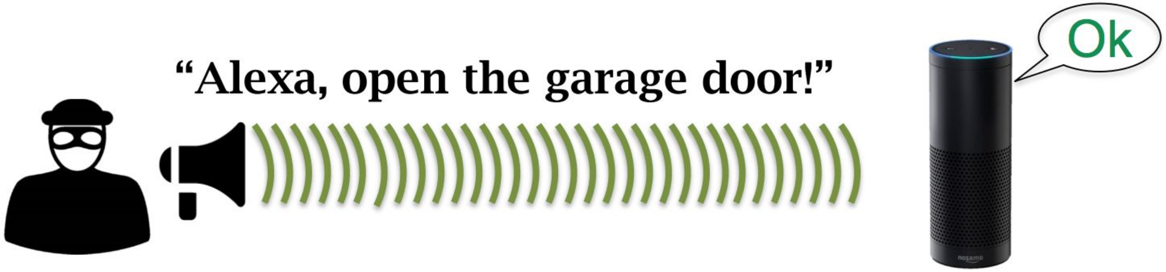
Liwei Song, Prateek Mittal
Department of Electrical Engineering,
Princeton University



Normal Sound
(< 24 kHz)

Ultrasound
(> 25 kHz)

“Inaudible Acoustics”
(> 25 kHz)



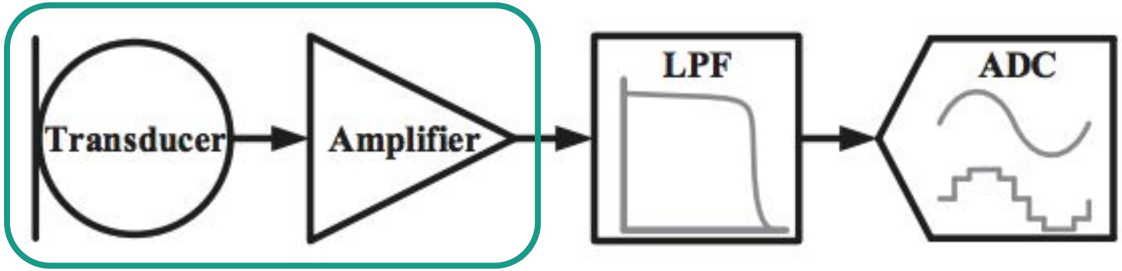
“Alexa, open the garage door!”

Agenda

- **Inaudible Voice Attack**
- How to increase the attack range?
- How to defend against these attacks?
- Evaluation

Inaudible Voice Attack

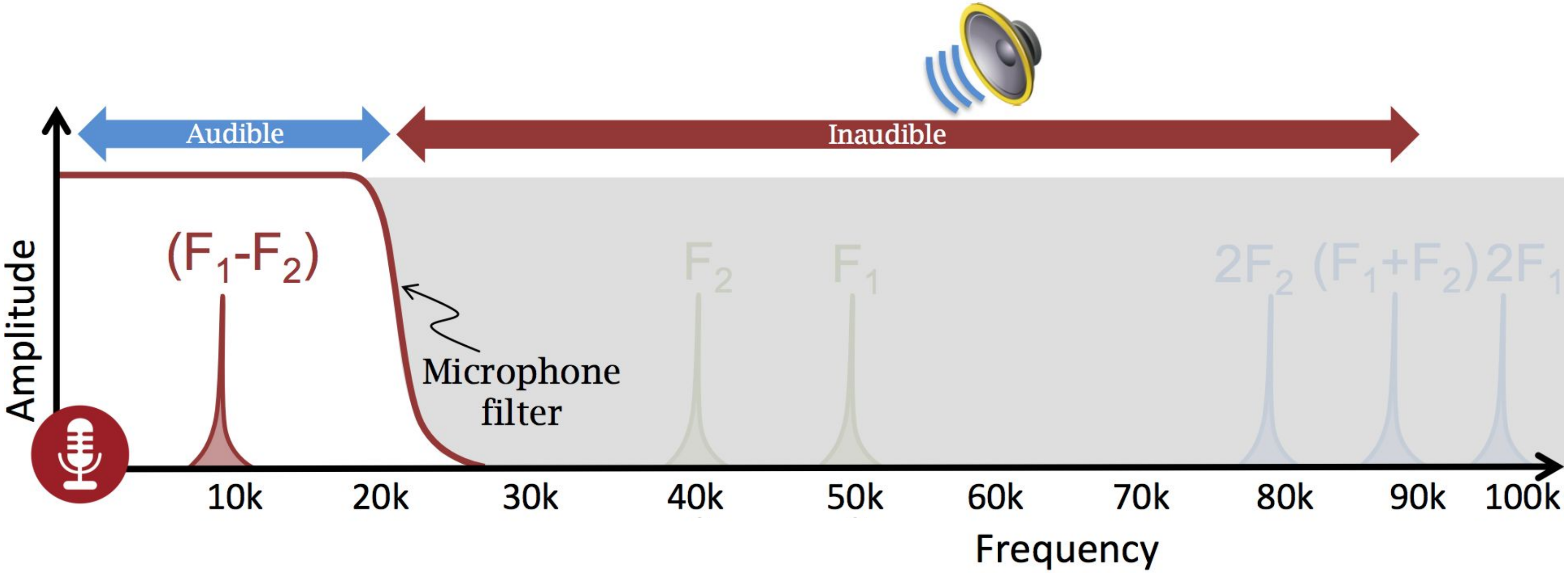
Exhibits non-linearity for ultrasound bands



$$s_{out}(t) \cancel{=} A_1 s(t)$$

$$\begin{aligned} s_{out}(t) &= \sum_{i=1}^{\infty} A_i s^i(t) = A_1 s(t) + A_2 s^2(t) + A_3 s^3(t) + \ldots \\ &\approx A_1 s(t) + A_2 s^2(t) \end{aligned}$$

Inaudible Voice Attack



Inaudible Voice Attack

BUT non-linearity requires high power . . .

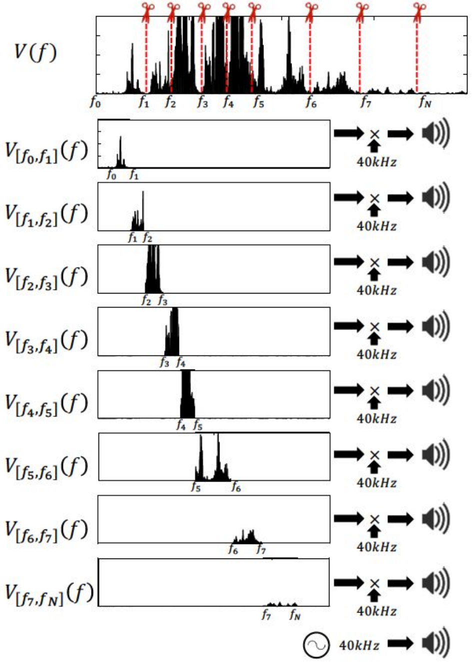
High power makes ultrasonic speakers audible

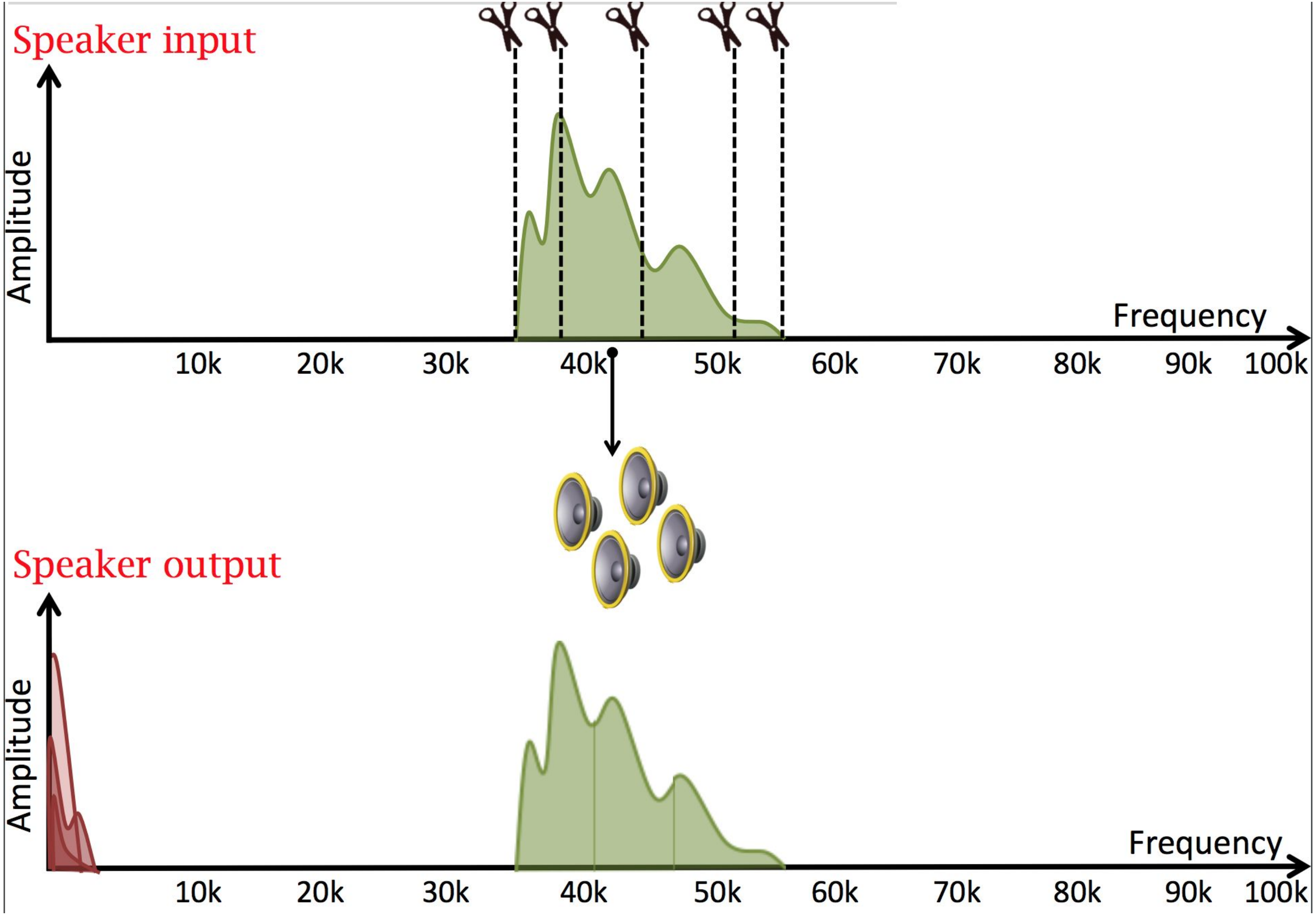
Agenda

- Inaudible Voice Attack
- **How to increase the attack range?**
- How to defend against these attacks?
- Evaluation

Long Range Attack

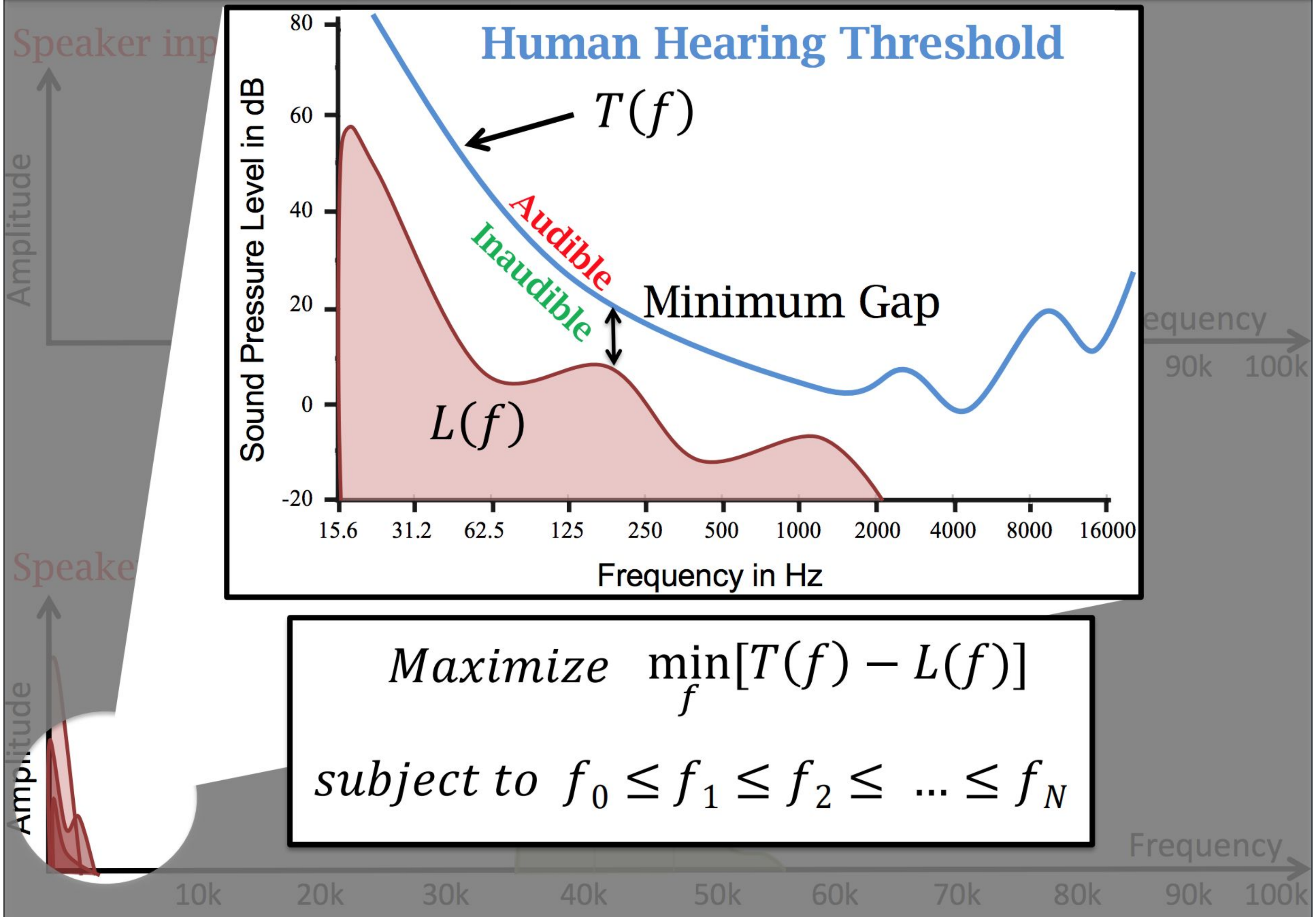
- Exploit non-linearity of speakers
- Use multiple ultrasound speakers to reduce the audible leakage from any speaker
  - Splice the input signal into segments
  - Each segment is played by a different speaker




Speaker input
Amplitude
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Speaker output
Amplitude
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k

Long Range Attack

- Per-speaker leakage is small but they can add up to become audible
- Ensure total leakage is inaudible
  - Push the total leakage below the threshold of hearing curve



$$Maximize \quad \min_{f}[T(f) - L(f)]$$

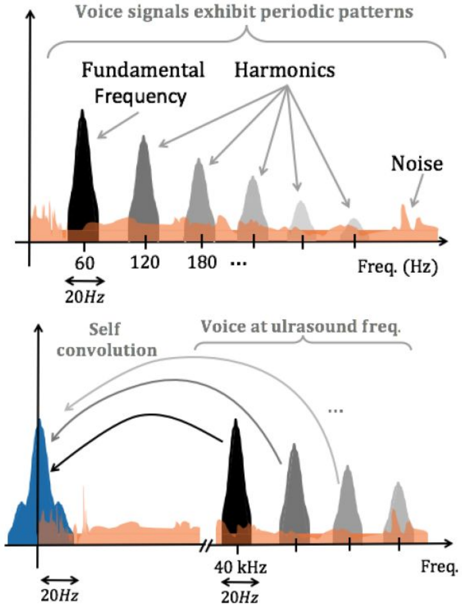
$$subject\ to \quad f_0 \leq f_1 \leq f_2 \leq \ \ldots \leq f_N$$

Agenda

- Inaudible Voice Attack
- How to increase the attack range?
- **How to defend against these attacks?**
- Evaluation

Defending Inaudible Voice Commands

- Find trace of non-linearity in the recorded signal
  - Power in sub-50 Hz
  - Correlation Coefficient
  - Amplitude Skew



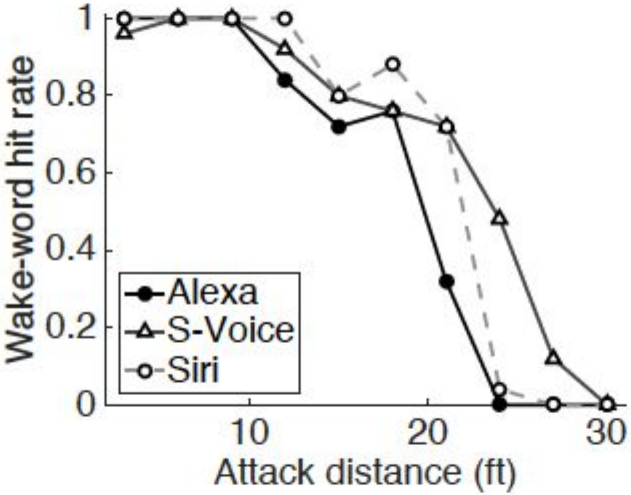
Defending Inaudible Voice Commands

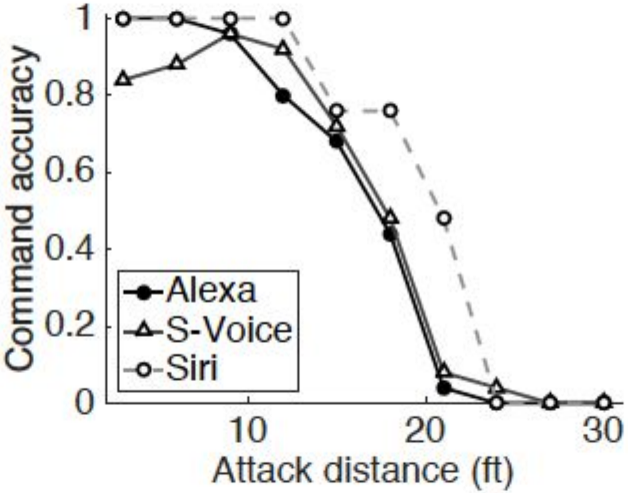
- Regular voice signals has less energy in sub-50Hz components
  - Non-linearity shows increasing power in that band
- Correlation between the power as non-linear trace preserves some structure of the actual input voice signal
- Amplitude of inaudible signal is positively biased

Agenda

- Inaudible Voice Attack
- How to increase the attack range?
- How to defend against these attacks?
- **Evaluation**

Attack Range Metrics





Discussion

- Feasibility of long range attack
  - Through-wall attack
- Other attacks on Voice Recognition System
  - Skill-Squatting
  - Other sounds

Other ideas ??