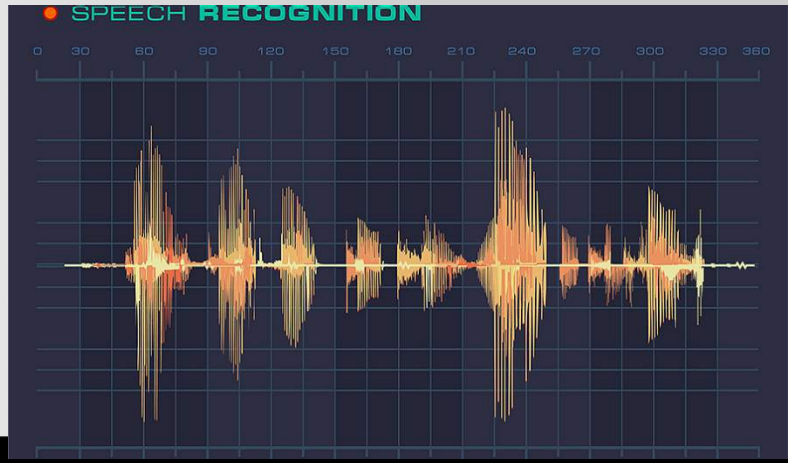
You Can Hear But You Cannot Steal
Defending against Voice Impersonation Attacks on Smartphones

Si Chen, Kui Ren, Sixu Piao, Cong Wang, Qian Wang,
Jian Weng, Lu Su, Aziz Mohaisen

Department of Computer Science and Engineering, University at Buffalo
Department of Computer Science, West Chester University
Department of Computer Science, City University of Hong Kong
Department of Computer Science, Wuhan University
Department of Computer Science, Jinan University



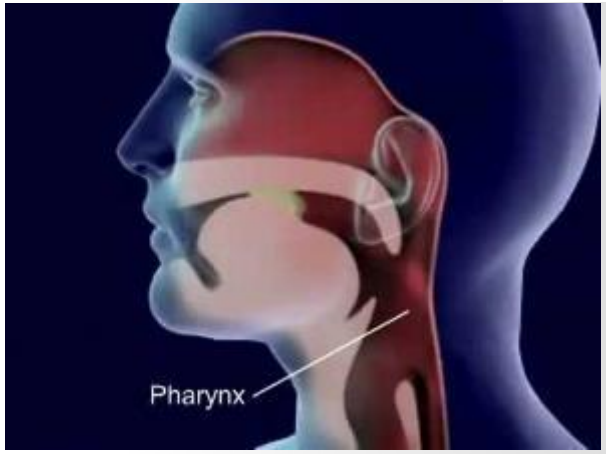
Outline

- **Introduction**
- Problem Formulation
- The Proposed Solution
- Evaluation
- Conclusion

Voice Authentication

Voice, has a significant advantage over the conventional keyboard-based input methods

- No memorization

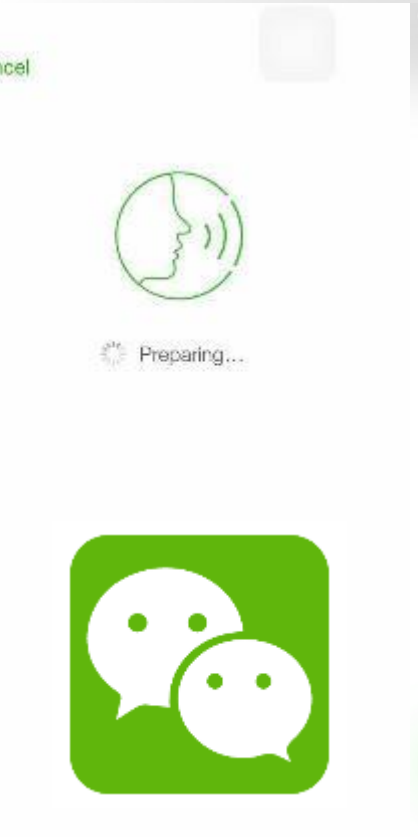


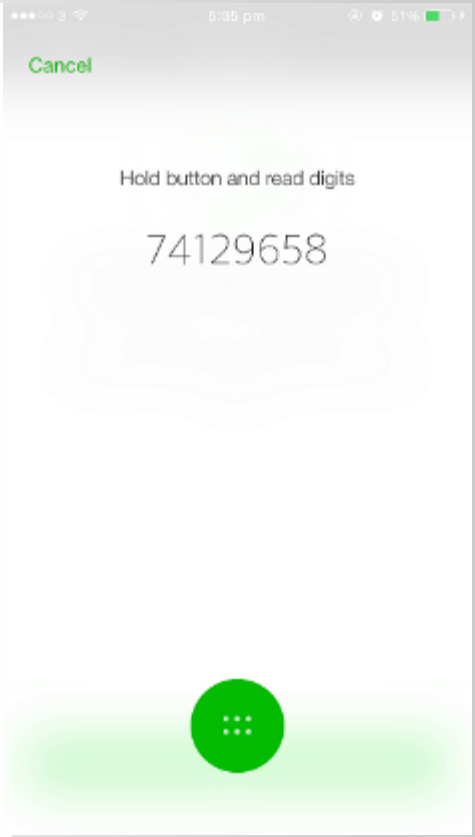
- Easy to Use

- Low Cost

Voice Authentication on Smartphones

Voice-enable Logins





E-Commerce

SayPay

Mobile Banking

HSBC offers voice and fingerprint ID system to customers

HSBC BARCLAYS citi

Outline

- Introduction
- **Problem Formulation**
- The Proposed Solution
- Evaluation
- Conclusion

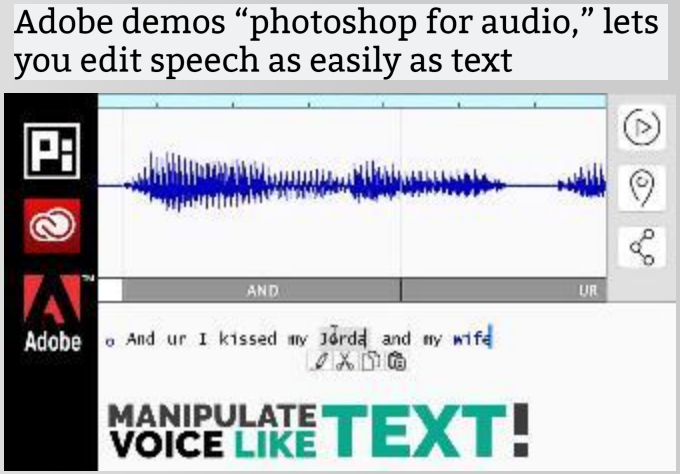
Problem Formulation

- The human voice could often be exposed to the public, an attacker can:
  - **Collect sound samples** of targeted victims
  - **Change voice biometrics** by using different methods
  - **Launch *voice impersonation attacks*** to spoof those voice-based applications

Voice Replay Attack

Voice Morphing Attack

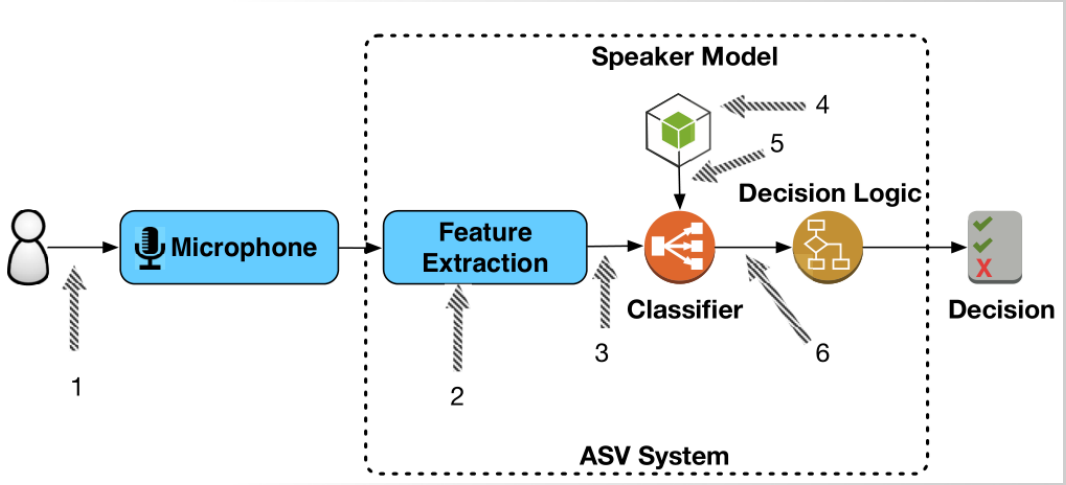
Voice Synthesize Attack



Adversary Model

- Voice impersonation attacks
  - **Machine-based Voice Impersonation Attack**
    - Voice Replay Attack
    - Voice Morphing Attack
    - Voice Synthesize Attack
  - **Human-based Voice Impersonation Attack**
    - Human Mimicking

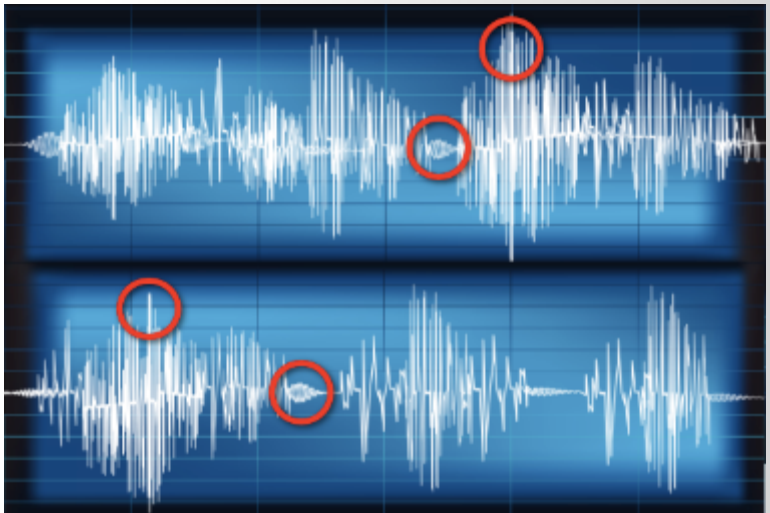
Existing Solutions: Automatic speaker verification (ASV) system



A generic automatic speaker verification (ASV) system with seven possible attack points

ASV system:

- Effective in detecting *human voice imitation* (human mimicking)
- Ineffective in detecting *machine-based voice impersonation* attacks



Problem Formulation

Can we build software-based defense system tailored for mobile platforms against voice impersonation attacks?

… and meet these design goals

- High accuracy
- Easy to integrate with off-the-shelf mobile phones
- Low latency
- Low computational cost

Problem Formulation

Can we build software-based defense system tailored for mobile platforms against voice impersonation attacks?

… and meet these design goals

- High accuracy
- Easy to integrate with off-the-shelf mobile phones
- Low latency

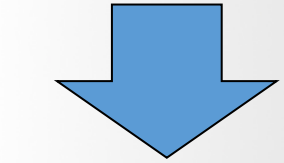
Outline

- Introduction
- Problem Formulation
- **Proposed Solution**
- Evaluation
- Conclusion

Proposed Solution

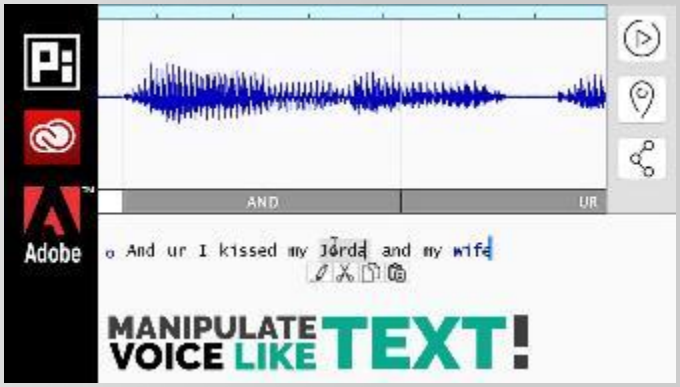
Voice Replay Attack

Voice Morphing Attack

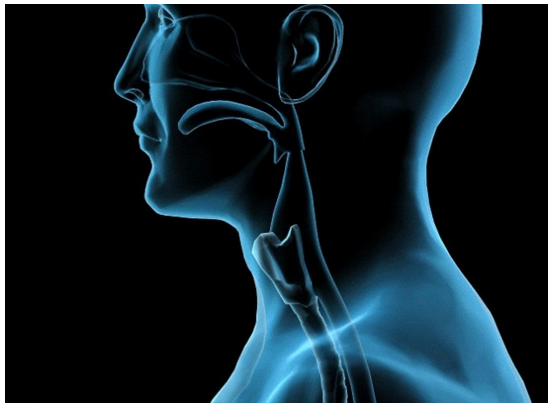


Voice Synthesize Attack

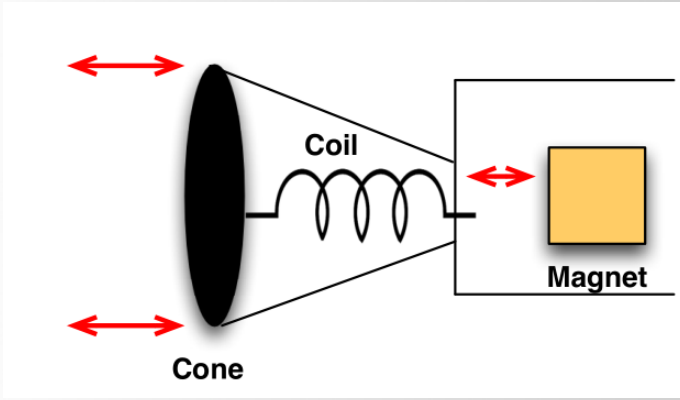
Adobe demos “photoshop for audio,” lets you edit speech as easily as text



Proposed Solution



The human vocal tract



The architecture of conventional loudspeaker

Key insight:

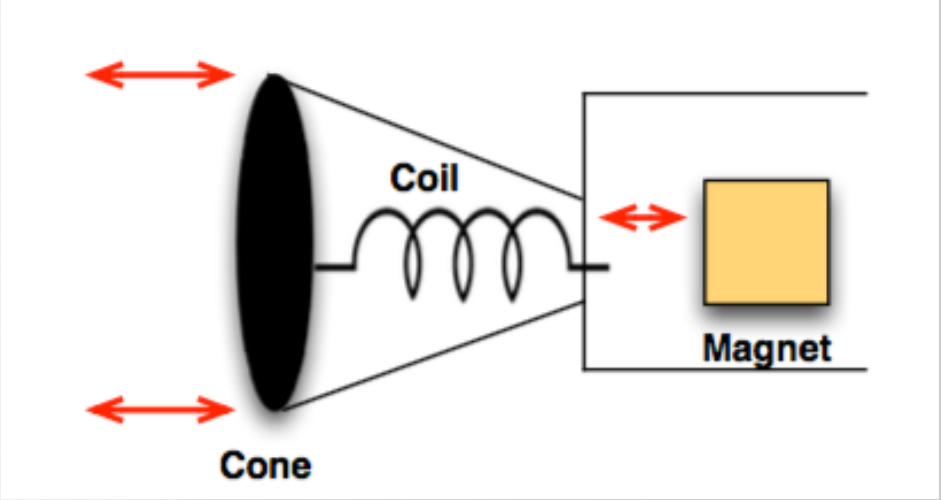
- The human vocal tract → No magnetic field
- The conventional loudspeakers → Has magnetic field

Use the magnetometer (compass) in smartphone to detect!

Proposed Solution

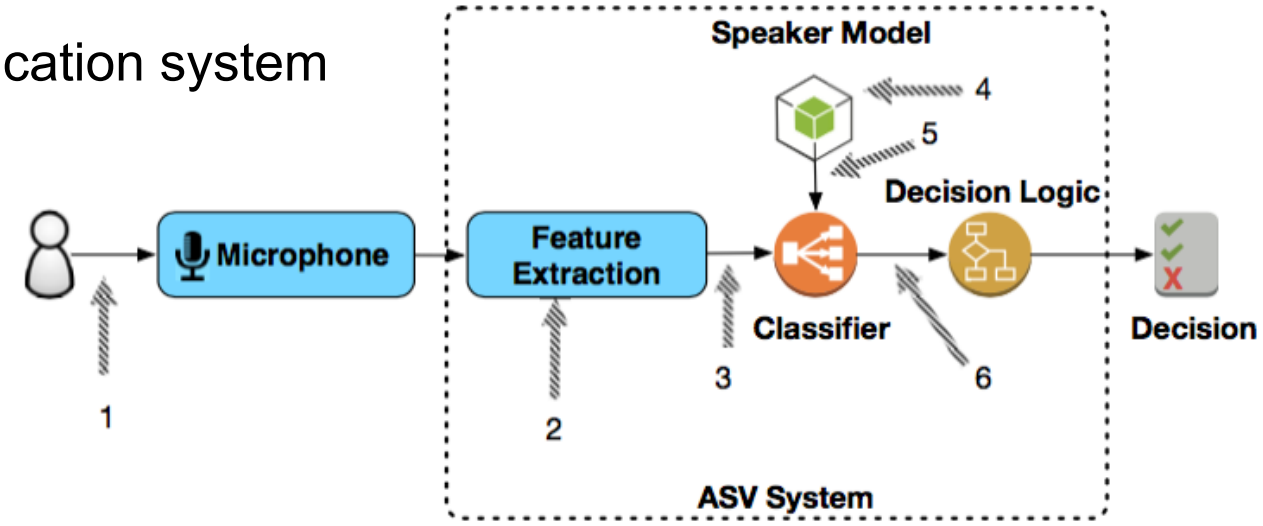
- **Machine-based Voice Impersonation Attack**

magnetometer (compass) equipped on modern smartphones

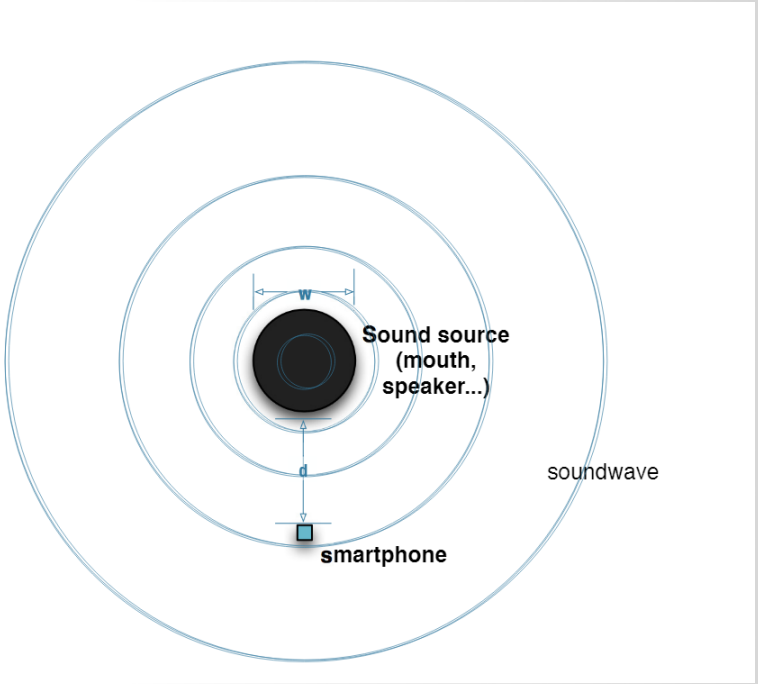


- **Human-based Voice Impersonation Attack**

Spear speaker verification system

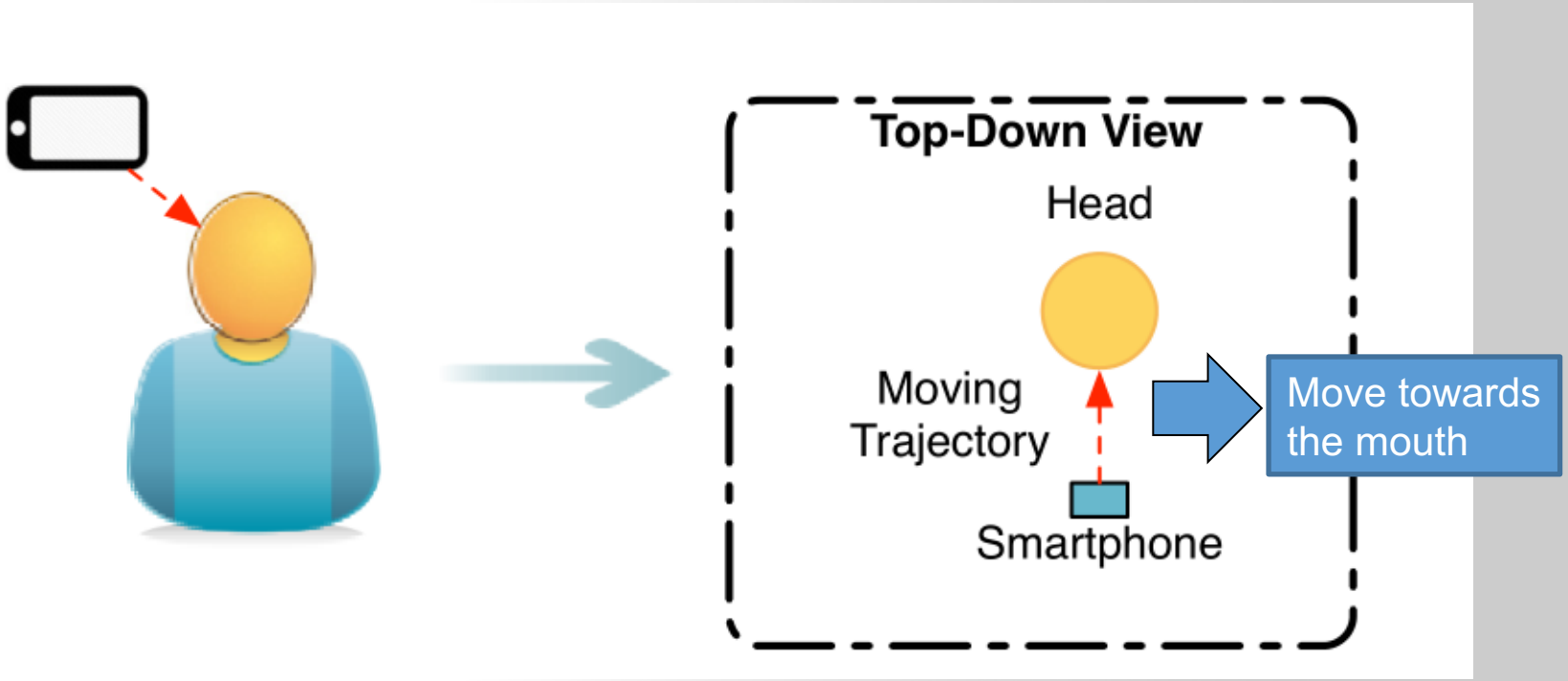


Proposed Solution

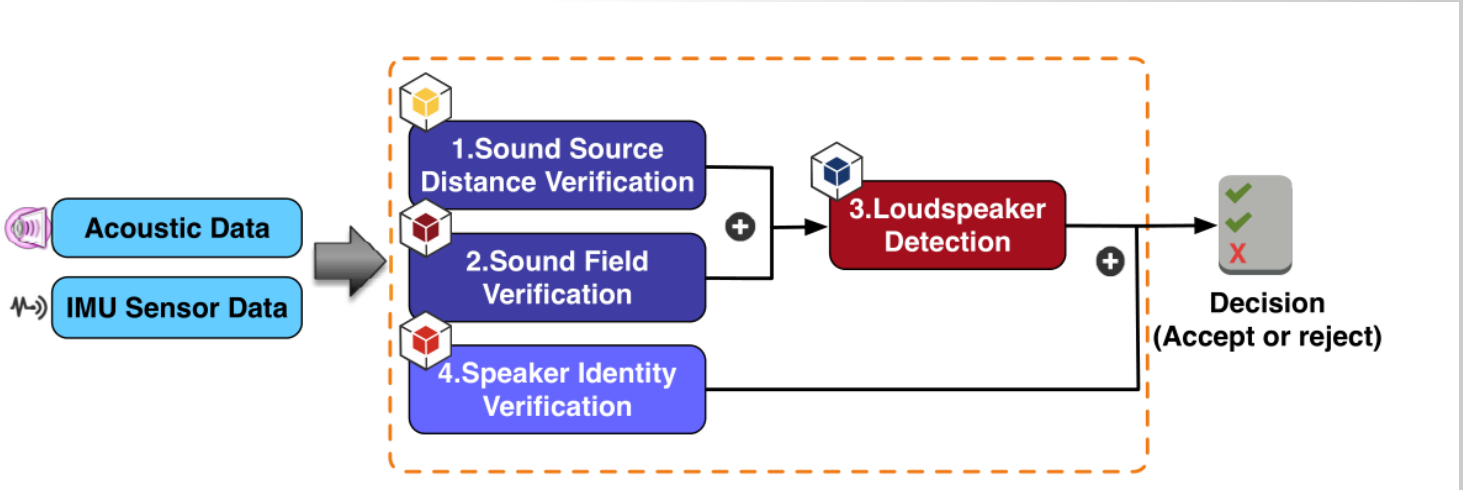


To successfully leverage our key insight, we require users to place the smartphone as close as possible to the sound source.

Proposed Solution



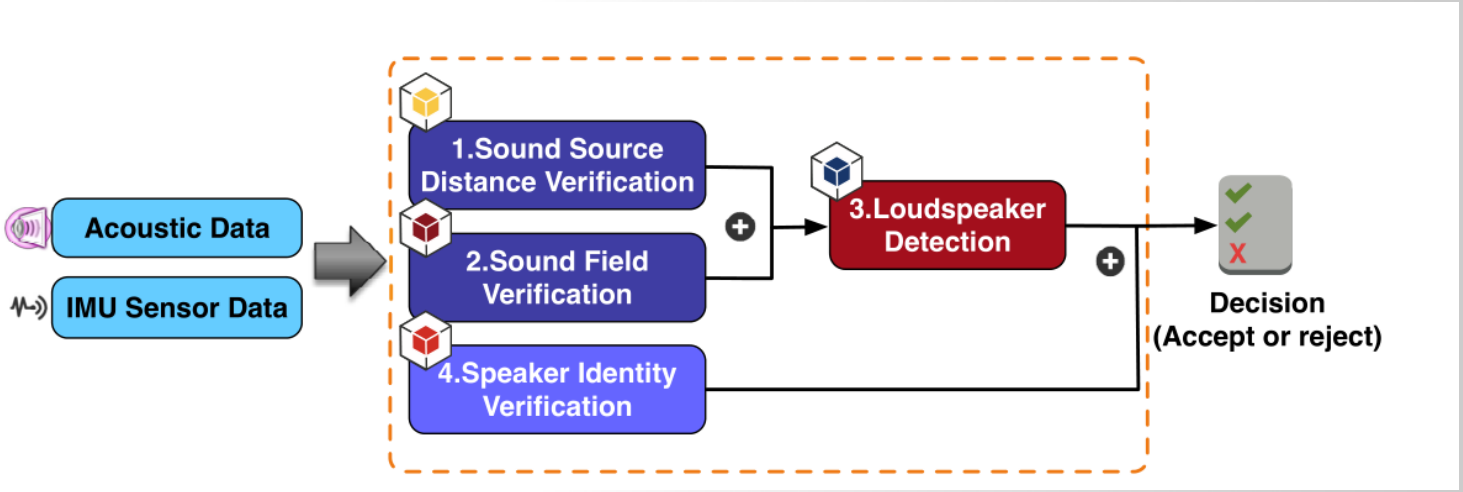
Proposed Solution



The architecture of our defense system

1. **Sound Source Distance Verification**
   - Reconstruct the moving trajectory of the smartphone
   - Calculate the distance between sound source and smartphone
2. **Sound Field Verification**
   - Justify whether the received sound is broadcast from a human mouth
3. **Loudspeaker Detection**
   - Detect the magnetic field emitted from the loudspeaker.
4. **Speaker Identity Verification**
   - Defend against human-based voice impersonation attacks

Proposed Solution



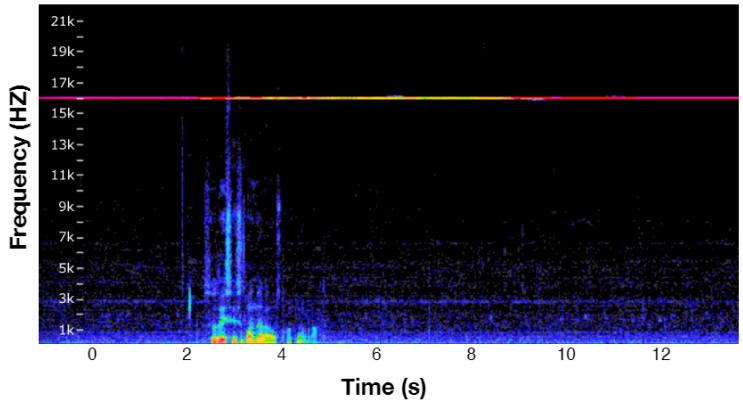
The architecture of our defense system

Our defense system consists of **four** verification components for defending against voice impersonation attacks:

- **Component 1, 2, 3:** Detect Machine-based voice impersonation attacks
- **Component 4:** Detect human-based voice impersonation attacks

The Proposed Solution

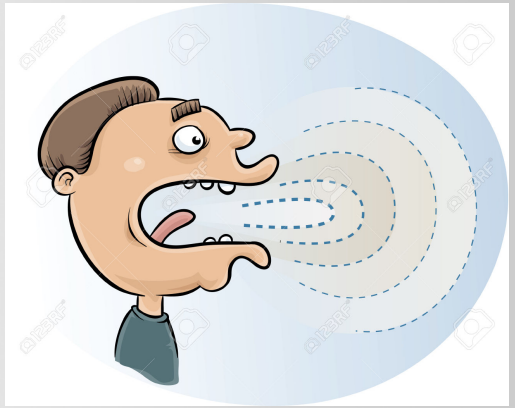
- ***Sound Source Distance Verification***
  - Reconstruct the moving trajectory of the smartphone
  - Calculate the distance between sound source and smartphone

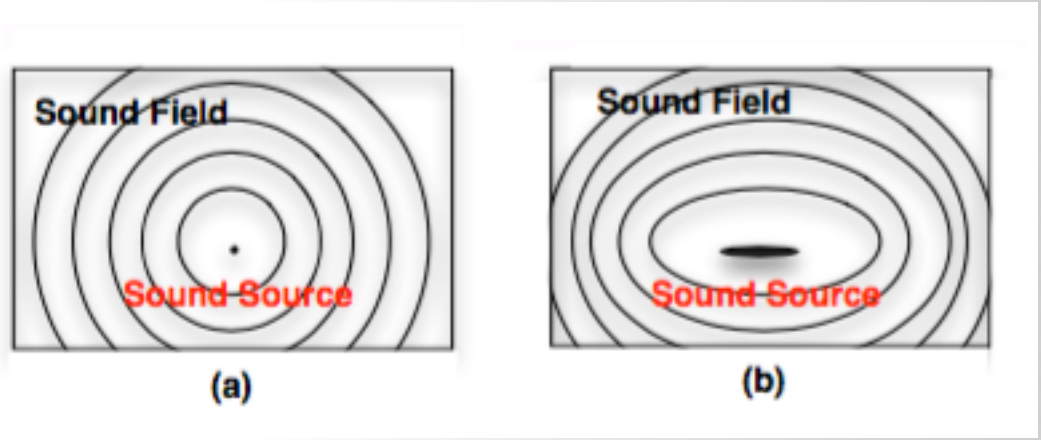


- Motion Trajectory Reconstruction
  - Acoustic sound
  - IMU Sensor

The Proposed Solution

- ***Sound Field Verification***
  - Justify whether the received sound is broadcast from a human mouth

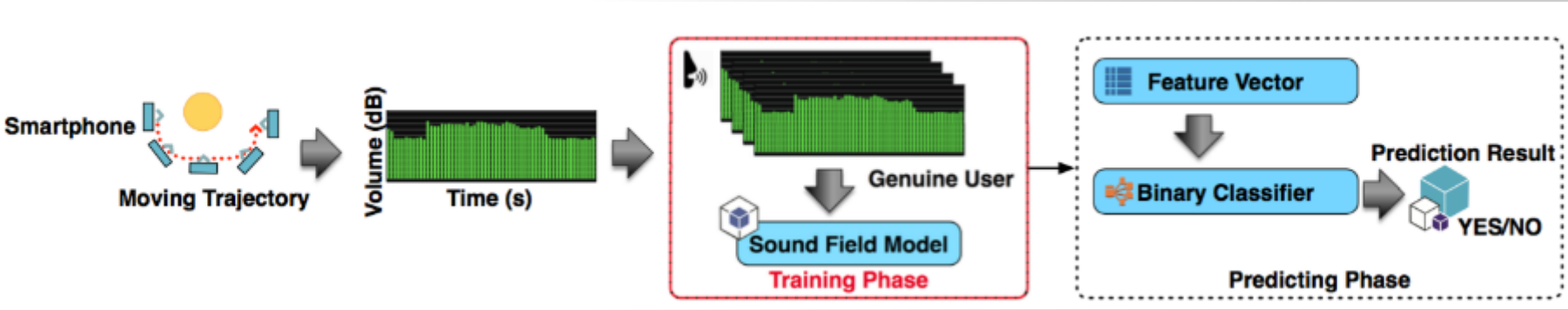




The sound field created by
(a) a point sound source and (b) created by a strip-type sound source.

The Proposed Solution

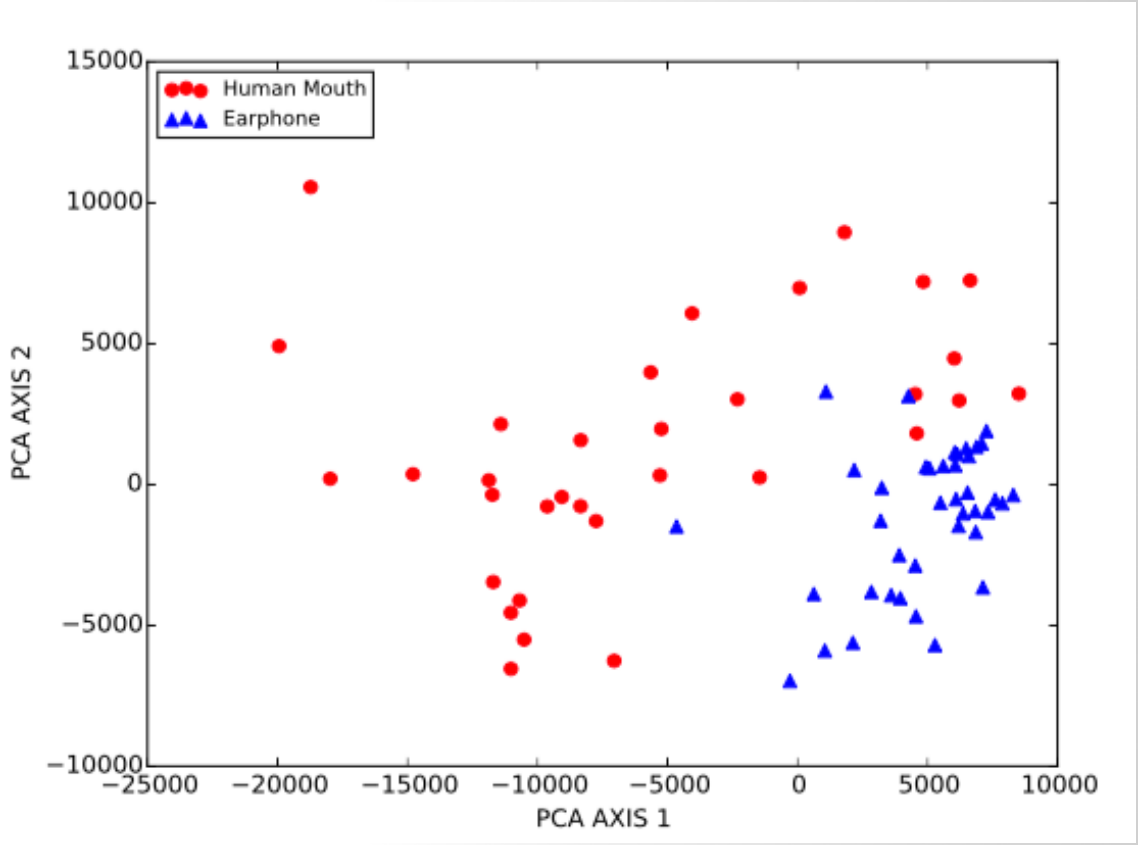
- ***Sound Field Verification***



The sound source validation process, containing two phases:
i) training phase and ii) predicting phase.

The Proposed Solution

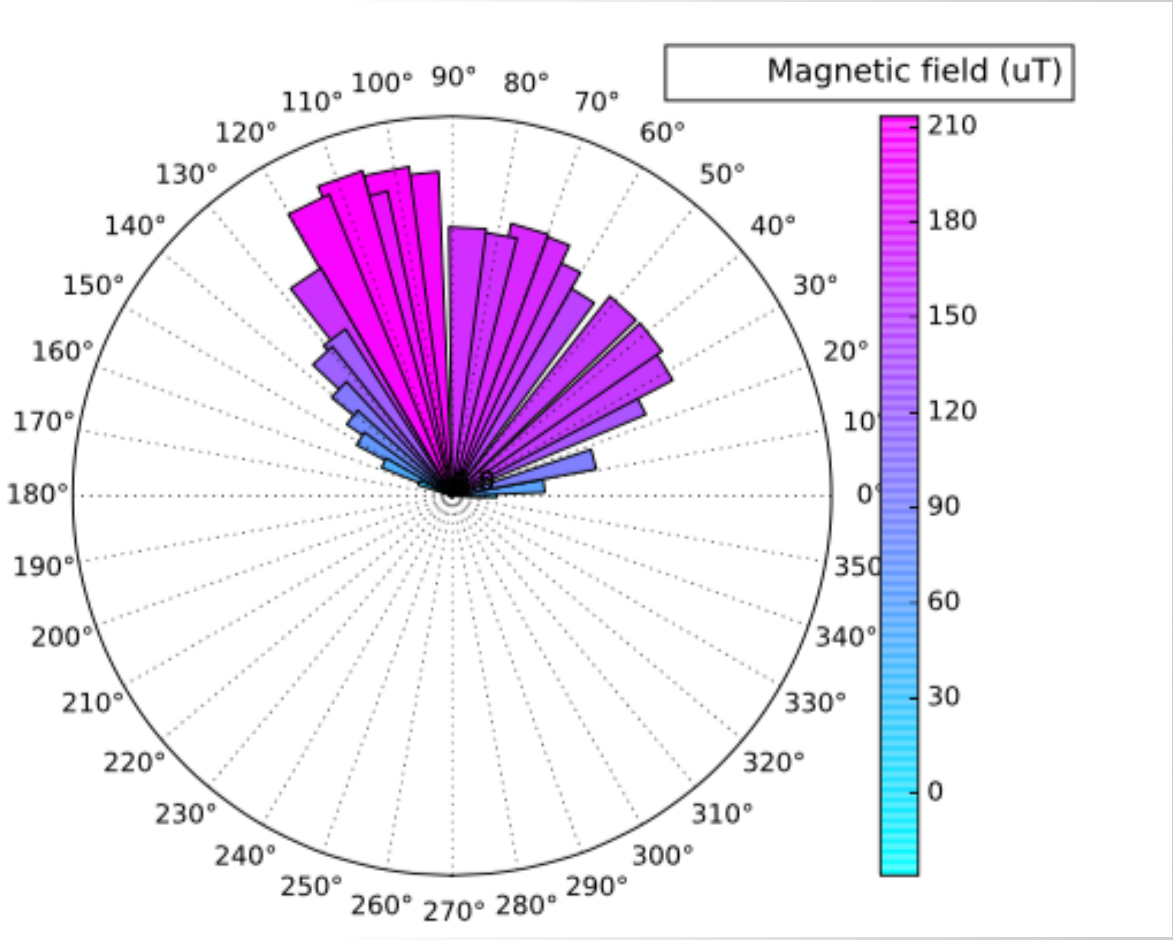
- ***Sound Field Verification***



The feature point of the human mouth sound field (red circle)
and the earphone sound field (blue triangle) after principal component analysis (PCA)

The Proposed Solution

- ***Loudspeaker Detection***
  - Detect the magnetic field emitted from the loudspeaker.



Polar graph of the magnetic field reading for a conventional loudspeaker.

Outline

- Introduction
- Problem Formulation
- The Proposed Solution
- **Evaluation**
- Conclusion

Evaluation

Methodology

- We design and build a small testbed environment
  - a real loudspeaker
  - a smartphone hardware.

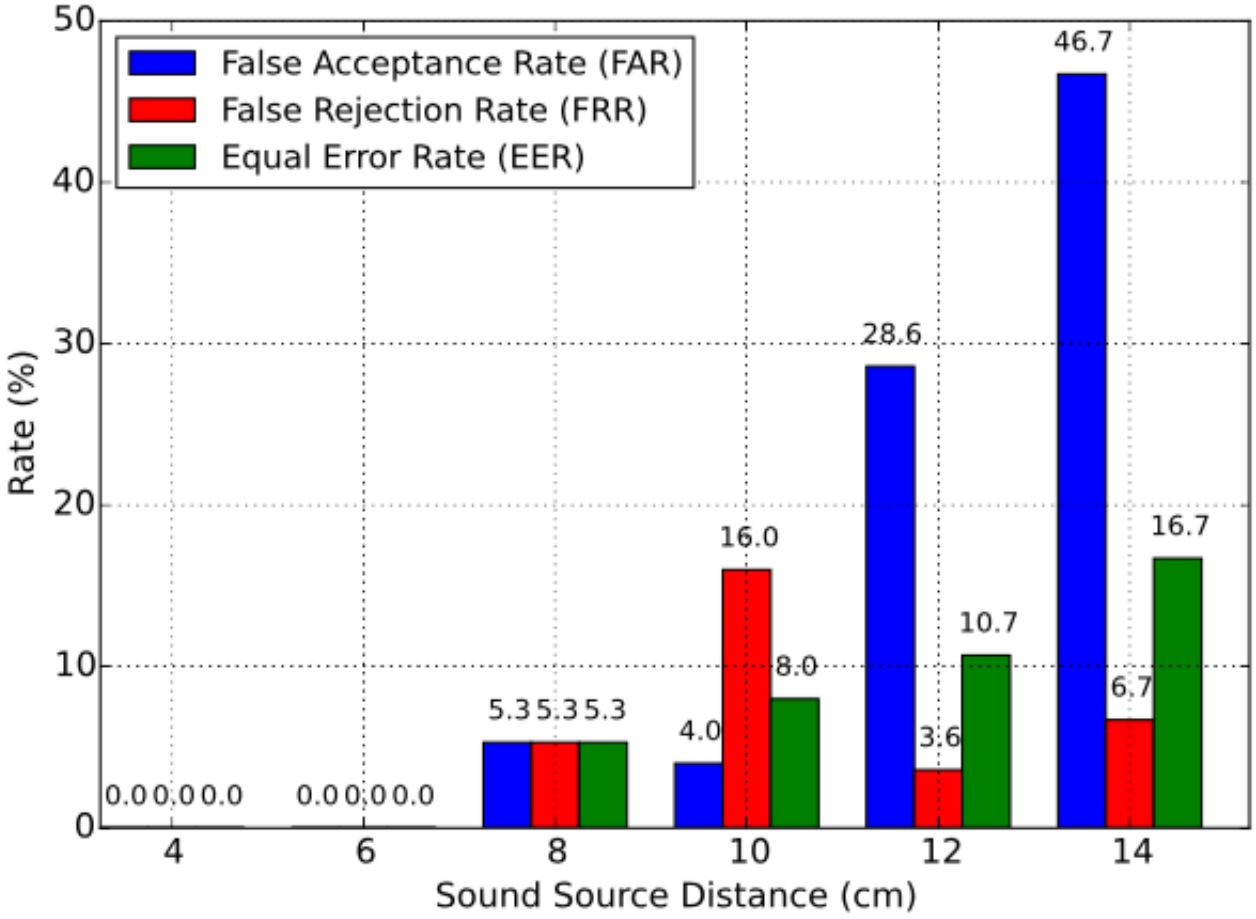
Our evaluation focuses on the machine-based voice impersonation anti-spoofing sub-system

Performance Metrics

- We choose the standard automatic speaker verification metrics
  - False Acceptance Rate (FAR)
  - False Rejection Rate (FRR)
  - Equal Error Rate (EER)
    - the rate at which the acceptance and rejection errors are identical

	Decision	
	Accept	**Reject**
Genuine	Correct Acceptance	False Rejction
Impostor	False Acceptance	Correct Rejection

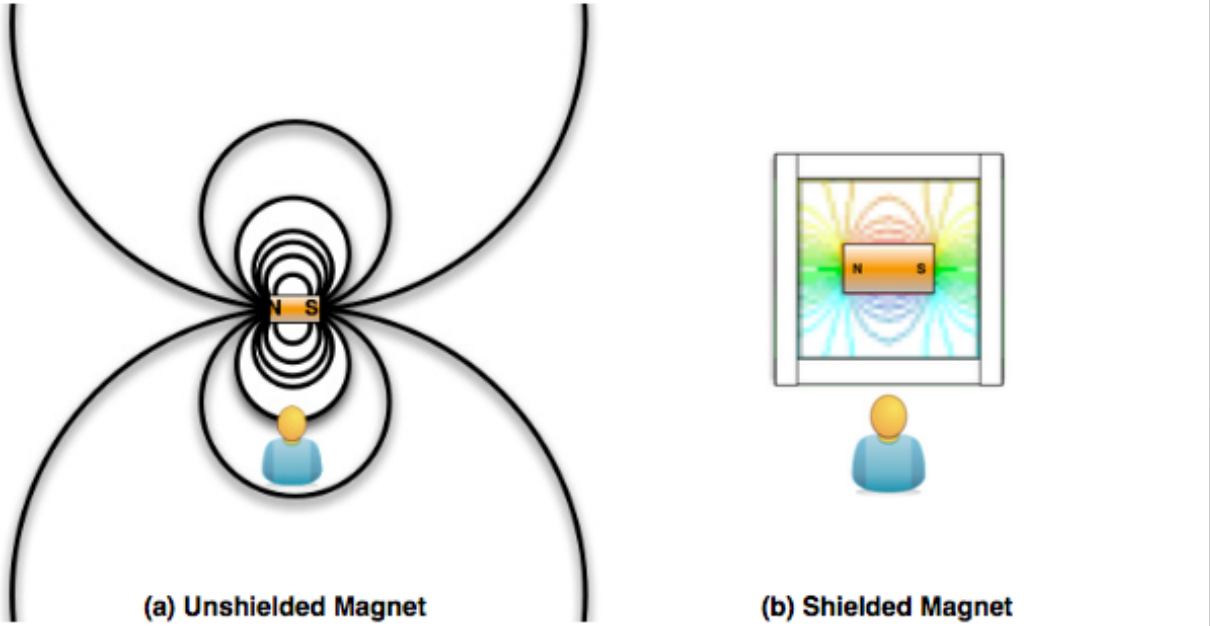
Evaluation



Impact of sound source distance of our defense scheme.

The FAR, FRR and EER values of our system are all equal to zero when the distance is less than or equal to 6 *cm*.

Evaluation



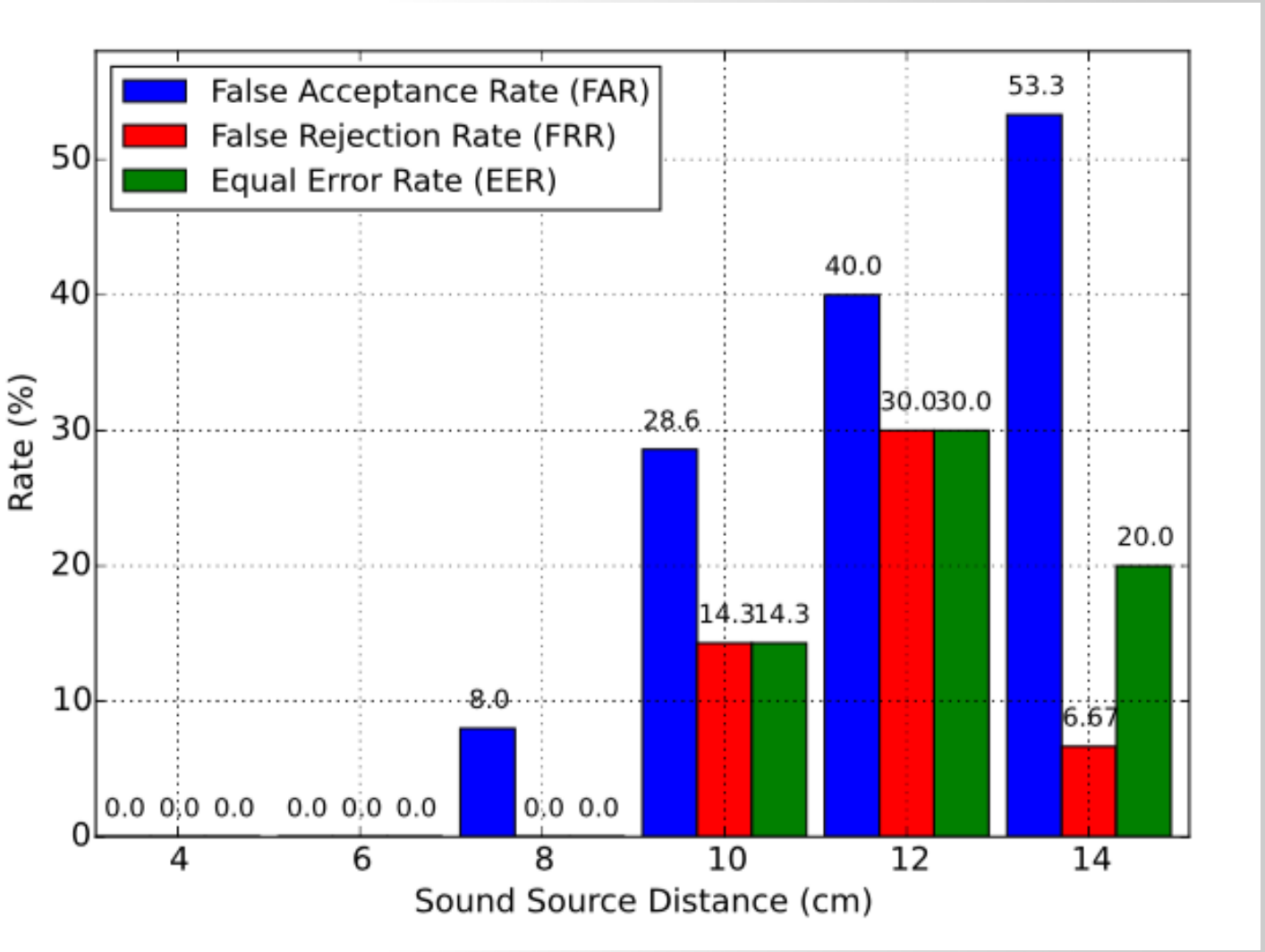
The magnetic field distribution of: (a) unshielded magnet and (b) shielded magnet.

Mu-metal

Mu-metal is a nickel-iron alloy

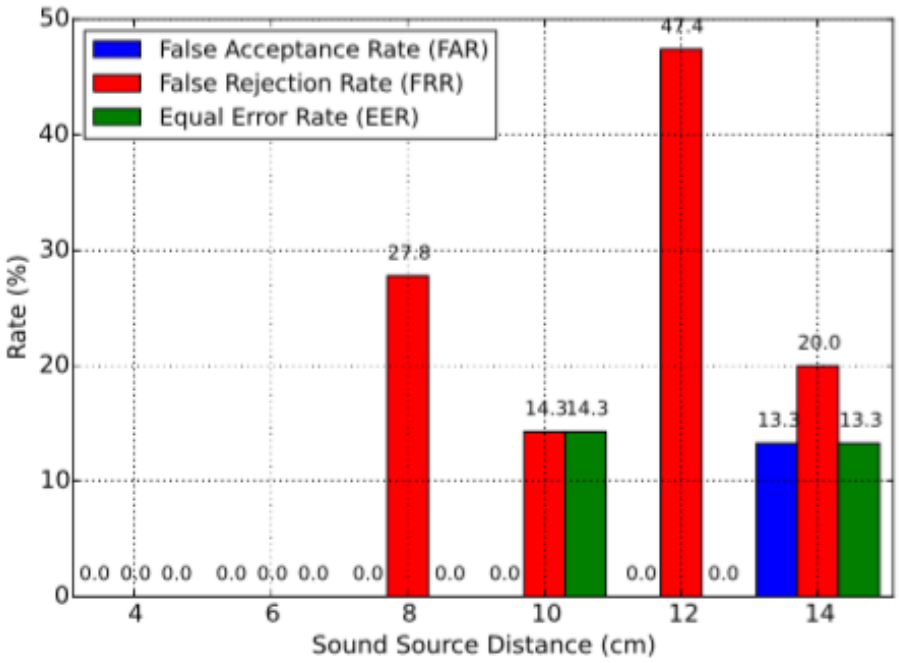
Perfect to shield the magnetic field.

Evaluation

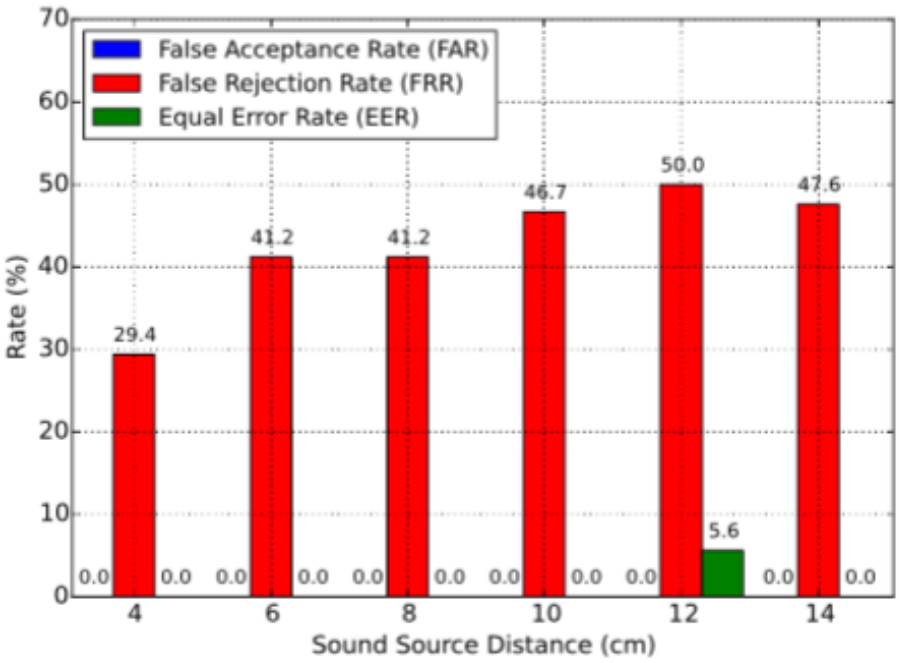


Impact of sound source distance for Magnetic field shielding of our defense scheme.

Evaluation



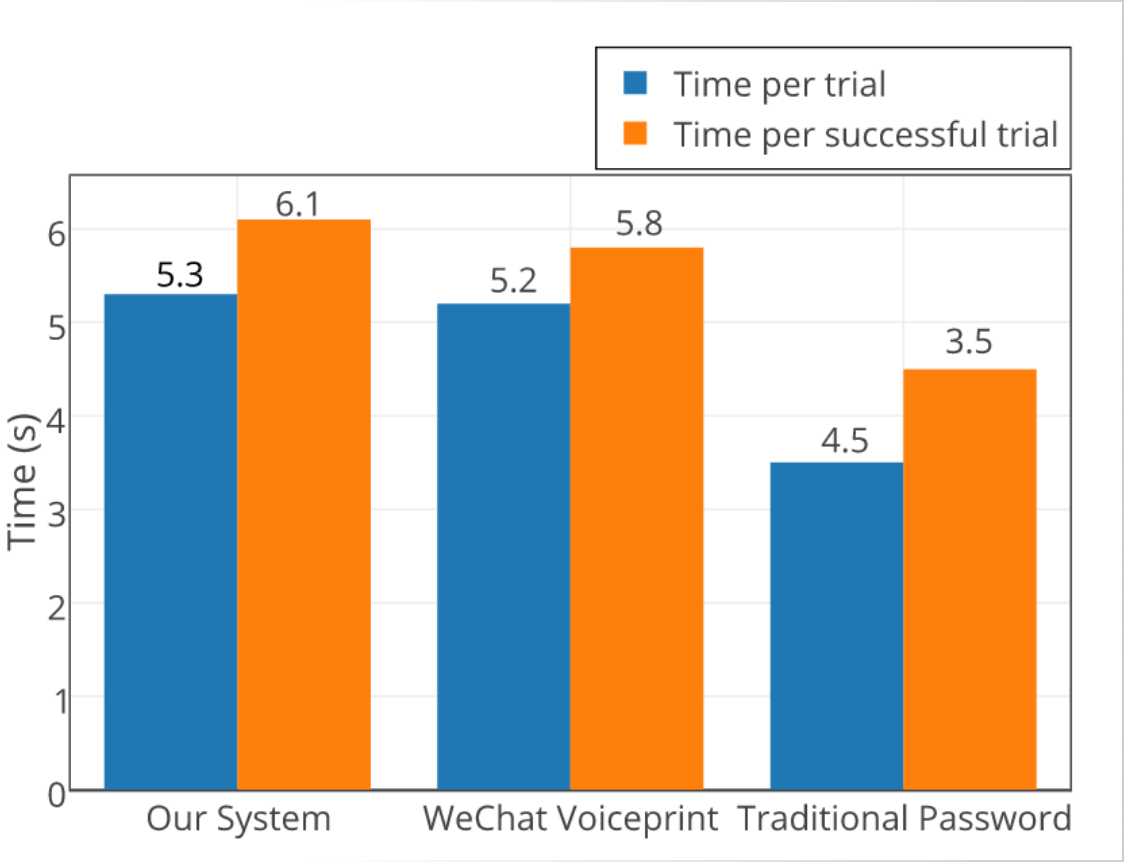
(a) Near a computer



(b) In a car.

The FAR, FRR and EER values of our system with environmental magnetic interference: (a) Near a computer (iMac 27' Late 2009) and (b) In a car's front seat (Hyundai Sonata 2012).

Evaluation



Authentication time comparison

Outline

- Introduction
- Problem Formulation
- The Proposed Solution
- Evaluation
- **Conclusion**

Conclusion

- Software-based solution tailored for mobile platform for defending against voice impersonation attacks
- Defeat the vast majority of voice impersonation attacks and **significantly raise the level of security** for existing voice-based mobile applications

- Our system achieves design goals
  - High accuracy (~100% accuracy when <= 6 cm)
  - Easy to integrate with off-the-shelf mobile phones (software-based approach)
  - Low latency (~ 6.1s for authentication)

Q & A