# 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, SUNY
[‡]Department of Computer Science, West Chester University
[¶]Department of Computer Science, City University of Hong Kong
[§]School of Computer Science, Wuhan University
[♮]School of Information Science and Technology, Jinan University
Email: schen@wcupa.edu, kuiren@buffalo.edu, sixupiao@buffalo.edu, congwang@cityu.edu.hk, qianwang@whu.edu.cn, cryptjweng@gmail.com, lusu@buffalo.edu, mohaisen@buffalo.edu

***Abstract*—Voice, as a convenient and efficient way of information delivery, has a significant advantage over the conventional keyboard-based input methods, especially on small mobile devices such as smartphones and smartwatches. However, the human voice could often be exposed to the public, which allows an attacker to quickly collect sound samples of targeted victims and further launch voice impersonation attacks to spoof those voice-based applications. In this paper, we propose the design and implementation of a robust software-only voice impersonation defense system, which is tailored for mobile platforms and can be easily integrated with existing off-the-shelf smart devices. In our system, we explore magnetic field emitted from loudspeakers as the essential characteristic for detecting machine-based voice impersonation attacks. Furthermore, we use a state-of-the-art automatic speaker verification system to defend against human imitation attacks. Finally, our evaluation results show that our system achieves simultaneously high accuracy (100%) and low equal error rates (EERs) (0%) in detecting the machine-based voice impersonation attack on smartphones.**

## I. Introduction

The proliferation of smartphones and wearable devices have fostered the booming of voice-based mobile applications [24], [33], which use human voice as a convenient and non-intrusive way for communication and command control. Common functionalities of these applications include traditional voice over IP (VoIP) (e.g., Skype and Hangouts), trending voice-based instant messaging (e.g., WeChat, TalkBox, and Skout), and intelligent digital personal assistant (e.g., Amazon Alexa, Google Home, Apple's Siri).

Even for security, voice has also been widely used in many mobile applications [51], [8] as a convenient and reliable way of user authentication. For example, WeChat provides "Voiceprint" [51], an authentication interface that allows users to log into WeChat by speaking pass-phrases. Baidu, a major Chinese web services company, also introduced voice-unlock as a built-in authentication method in their smartphone operating system [8]. With the exploding market of smart mobile devices, the voice-based mobile applications are expected to become even more popular in the next few years [33].

However, unlike other human biometrics, the human voice could often be exposed to the public. Examples of such exposure include scenarios where people are present in public receiving phone calls, or just talking loud in a restaurant. As such, an attacker could easily "steal" a victim's voice by just using handy recorders such as smartphones, by downloading the audio clips from the victim's online social networking website [7], or even by creating and recording a spam call. Upon the successful collection of enough voice samples, a high fidelity acoustic model of the victim's voice can be then reconstructed with the current advancement in voice processing [25]. Using the victim's acoustic model, an adversary could easily convert his voice into the victim's voice using voice morphing techniques. With state-of-the-art speech synthesis techniques (e.g. Adobe Voco [36]), even synthetic speech that resembles the victim's voice could be generated using any provided text.

Because voice is commonly characterized as one of the unique biometric features for personal authentication [13], an adversary that can imitate the victim's voice would quickly launch voice impersonation attacks to spoof any voice-based applications [53], [39]. This, in turn, would result in severe consequences to harm victim's reputation, safety, and property. For example, by spoofing the voice-based authentication mechanism, the attacker could easily steal private information from the victim's smartphone. Furthermore, fake voice calls or scam voice messages could be used to fraud the victim's social contacts.

The traditional methods of defending against the voice impersonation attacks require an automatic speaker verification (ASV) system, which employs unique spectral and prosodic features of a user's voice for user authentication [2], [40]. However, current ASV systems are far from perfect. While they are effective in detecting human-based voice impersonation attacks (human voice imitation) [5], [9], they are widely known for their inability to detect voice replay attacks [53]. Moreover, when detecting voice impersonation attacks, current

ASV systems require a prior knowledge of specific voice impersonation techniques used by the attacker [29]. Such an assumption does not necessarily always hold in practice. For example, one recent work [54] has demonestrated that ASV alone could be subject to sophisticated machine-based voice attacks. Hence, more robust designs resilient to both human-based and machine-based voice impersonation attacks are in great demand yet to be fully explored.

To build a robust defense system, there are many challenging barriers to overcome. One of the critical challenges is to defend against both human-based and machine-based attacks simultaneously. To achieve this goal, we leverage the following insights: in machine-based voice impersonation attacks (such as the replay attack, voice morphing attack, and voice synthesize attack), an attacker usually needs to use a loudspeaker (e.g., PC loudspeaker, smartphone loudspeaker, and earphone) to transform the digital or analog signal into the sound. The conventional loudspeaker uses magnetic force to broadcast the sound and leads to the generation of a magnetic field. Thus, if we can capture this magnetic field by monitoring the magnetometer reading from the smartphone, we can leverage it as a key differentiating factor between a human speaker and a loudspeaker. By carefully integrating our detection method with the current AVS systems, we can achieve a much more robust design to defend against all types of voice impersonation attacks on smartphones.

In addition to defending against attacks launched via conventional loudspeakers, we also consider special cases of machine-based voice impersonation attacks launched via small earphones. In such scenarios, the magnetic force emitted can be too small to be sensed directly by the magnetometer. To address this challenge, we resort to detecting the channel size of the sound source, and design a sound field validation mechanism to ensure that the sound source size is always close to a human mouth (i.e., not an earphone). By cross-checking both approaches, together with the careful integration of an existing AVS system, we can defeat the vast majority of voice impersonation attacks and significantly raise the level of security for existing voice-based mobile applications.

**Contribution.** Our main contributions are as follows:

1) We propose a robust software-only defense system against voice impersonation attacks, which is tailored for mobile platforms and can be easily integrated with off-the-shelf mobile phones and systems.

2) We use advanced acoustic signal processing, mobile sensing, and machine learning techniques, and integrate them as a whole system to efficiently detect voice impersonation attacks.

3) We build our system prototype and conduct comprehensive evaluations. The experimental results show that our system is robust and achieves very high accuracy with zero equal error rates (EER) in defending against voice impersonation.

**Organization.** In the rest of the paper, we begin with the background and related work in Section II, followed by the problem formulation in Section III. Section IV describes the scheme overview and design details. The implementation details are presented in Section V. The evaluation results are in Section VI. We further discuss our solution in Section VII. Finally, Section VIII concludes this paper.

## II. Background and Related Work

**Voice-based Mobile Applications.** Based on their functionality, existing voice-based mobile applications can be divided into two categories: i) *voice communication* ii) *voice control*. For voice communication, there are VoIP apps and instant voice message apps. As previously stated, by imitating a victim's voice, tone and speaking style, the attacker could easily launch impersonation attacks that would lead to severe harm to the victim. On the other hand, the applications in the second category allow users to use their voice commands to control the smartphone, using services such as the *voice recognition and assistant* and *voice authentication*. For voice recognition and assistant, Siri and Google Voice Search (GVS) are two noteworthy representative systems on iOS and Android systems, respectively.

In [14], the authors presented a recent threat that uses GVS application to launch voice-based permission bypassing attack and steal private user information from smartphones. As for voice authentication, quite a few mobile apps have adopted it as a built-in method for user authentication and system login. Besides the aforementioned WeChat "Voiceprint" [51] interface, Superlock [20] is another example that utilizes user's voice to lock and unlocks the phone. Unfortunately, a recent study shows that these authentication systems could be spoofed by an attacker mimicking the voice of the victim [53].

**Automatic Speaker Verification (ASV) System.** An ASV system can accept or reject a speech sample submitted by a user, and verify her as either a genuine speaker or an imposter [43], [27]. It can be *text-dependent* (with required utterances from speakers) or *text-independent* (able to accept arbitrary utterances) [10]. Text-independent ASV systems are more flexible and are able to accept arbitrary utterances, i.e., different languages, from speakers [10]. The text-dependent ASV is more widely selected for authentication applications, since it provides higher recognition accuracy with fewer required utterances for verification. The current practice for building an ASV system involves two processes: offline training and runtime verification. During the offline training, the ASV system uses speech samples provided by the genuine speaker to extract certain spectral, prosodic (see [2] and [40]) or other high-level features (c.f. [15] and [35]), to create a speaker model. Later in the runtime verification phase, the incoming voice is verified against the trained speaker model.

As shown in Fig. 1, a generic ASV system contains seven vulnerability points. Attacks at point (1) are the *voice impersonation attacks*, where the attacker tries to impersonate another person by using pre-recorded or synthesized voice sample before transmitting them into the microphone [23]. Attacks at point (2-6) are the *indirection attacks* [32], which are performed within the ASV system. In our paper, we build our defense system focusing on the first type of the attacks.

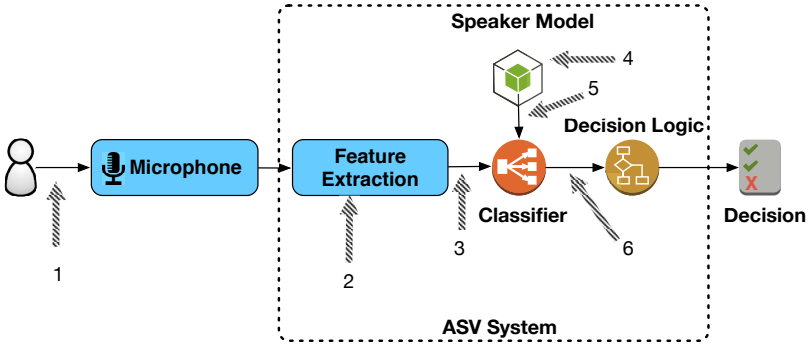



Fig. 1: A generic automatic speaker verification (ASV) system with seven possible attack points. The attack at point 1 denotes the voice impersonation attacks, whereas the attack at points 2 through 6 denote the indirection attacks.

**Voice Impersonation Attack.** The voice impersonation attack implies an attack targeting the ASV system using a pre-recorded, manipulated or synthesized voice samples to deceive the system into verifying a claimed identity [29]. The work in [26] suggests that, even though professional human impersonators are more effective than the untrained, they are still unable to repeatedly fool an ASV system. To address the human-based voice impersonation attacks, the work in [5], [9] proposed a disguise detection scheme. The scheme exploits the fact that voice samples submitted by an impersonator are less practiced and exhibit larger acoustic parameter variations. In particular, [5] claims a 95.8% to 100% detection rate for human-based impersonation attacks.

Another method of voice impersonation is the *machine-based voice impersonation attack*, such as replay attack, voice synthesis or conversion attack. To launch this type of attack, the attacker needs to seek help with specific devices (e.g., microphone, computer and loudspeaker). In [46], the author shows that an attacker can concatenate speech samples from multiple short voice segments of the target speaker and overcome text-dependent ASV systems by launching replay attacks. Although a few system research papers on developing replay attack countermeasures have been published [30], [38], [46], [47], [50], all these systems suffer from high false acceptance rate (FAR) compared to the respective baselines. In [4], the authors demonstrate vulnerabilities of ASV systems for voice synthesis attacks with artificial speech generated from text input. The work in [42], [55] propose the voice conversion attack in which the attacker converts the spectral and prosody features of her own speech in resembling the victim's. To detect voice synthesis and voice conversion attack, [56] exploited artifacts introduced by the vocoder to discriminate converted speech from original speech. A more recent work [3] claims a method that can detect voice conversion attack effectively by estimating dynamic speech variability.

The essential difference between our work and previous studies lies in the method we use for machine-based voice impersonation detection. We design a more general countermeasure by leveraging smartphone-equipped magnetometer to detect the magnetic field produced by the conventional loudspeakers. We then use this physical characteristic of the conventional loudspeakers to detect machine-based impersonator on smartphones, instead of analyzing the acoustic features of speech samples.

## III. Problem Formulation

### A. Adversary Model

The voice impersonation attack aims at attacking biometric identifiers of a system. In our adversary model, an attacker is able to collect the voice samples of the victim. As mentioned previously, this can be achieved by the attacker with little cost, since human voice could often be exposed to the public. Once an attacker acquires the voice samples, the attacker is able to use different methods to change their voice biometrics to appear like the victim. Then, the attacker can perform spoofed phone calls, or launch replay attacks, voice conversion attacks and voice synthesis attacks, through voice messaging and voice authentication applications. Based on the methods the attacker uses, we divide the voice impersonation attacks into the following two categories:

**1) Machine-based Voice Impersonation Attack.** In this type of attack, the attacker has the ability to leverage computer and other peripherals (e.g., loudspeaker) to gain the capability of voice replaying or voice morphing. Therefore, the attacker can imitate the target's voice at a high degree of similarity. We assume the attacker has a permanent or temporary access to the mobile application's front-end, which displays the voice-based I/O interface (e.g., a victim's mobile phone). Based on the capability of the attacker, we can further divide the machine-based voice impersonation attacks into three types.

**Type 1: Voice Replay Attack.** In this type of attack, the attacker is able to acquire an audio recording of the target's voice prior to the attack. The attacker tries to spoof the speaker verification system by replaying the voice sample using a loudspeaker.

**Type 2: Voice Morphing Attack.** In this type of attack, the attacker is able to imitate the target's voice by applying voice morphing (conversion) techniques. We assume that the voice spoofing techniques used by the attacker can produce high-quality output with all details of the human vocal tract. Moreover, the attacker has the ability to simulate the excitation

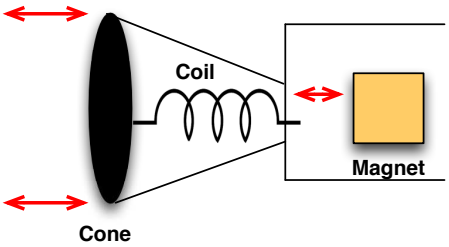



Fig. 2: The architecture of conventional loudspeaker showing the magnet, coil and cone used for loudspeaker operations.

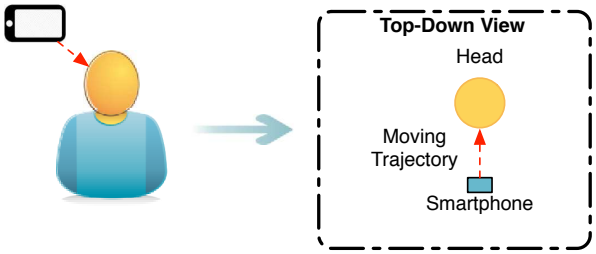



Fig. 3: A typical use case of our system.

of the vocal tract naturally. The attacker tries to spoof the speaker verification system by broadcasting the morphed voice using a loudspeaker to impersonate the targeted legitimate user.

**Type 3: Voice Synthesize Attack.** This type of attacker is able to synthesize target voice by using the state-of-the-art speech synthesizers techniques. We assume the attacker is able to use text-to-speech (TTS) technique to generate the natural-sounding synthetic speech of the targeted user from any input texts. The attacker tries to spoof the speaker verification system by directly broadcasting the synthetic voice using a loudspeaker.

We note that in the last step of each of the three types of attacks, a loudspeaker (e.g., PC loudspeaker, smartphone loudspeaker, etc.) is required to broadcast the processed voice. Thus, if the differentiation between the voice produced by a human and by a loudspeaker is clear, we can defend against the machine-based voice impersonation attacks from the source validation. The key insight of our design is discussed in the following section.

**2) Human-based Voice Impersonation Attack.** This type of attack, the attacker utilizes the acquired voice sample to imitate the target's voice without the help of any computer or professional devices. In particular, the attacker may use his voice or could seek help from other people (e.g., someone who can imitate the target's voice very closely). To defend against this type of attack, we utilize the state-of-the-art ASV system which leverages the acoustic features from the voice samples to perform voice impersonation attack detection.

### *B. Key Insights*

Our key goal is to differentiate genuine speakers from both machine-based and human-based impostors on smartphones. For human-based impostor, there already exist sophisticated speaker verification systems, such as the open-sourced Bob Spear verification toolbox developed by Khoury et al. [21], which has been recognized for its performance in detecting against human-based impersonation attacks [53], [5], [9].

For the machine-based impersonation attack, the existing state-of-the-art voice authentication systems can be easily circumvented by voice replay and conversion tools (e.g., Festvox [16]), among others. Therefore, relying on the spectral and prosodic features within the voice to defend against machine-based voice impersonation attacks has been proven ineffective. Thus, we address this problem from a new perspective.

We note that different from human-based voice impersonation, the machine-based impersonation attack requires the attacker to convert the digital signal to an audible sound by the assistance of a loudspeaker. Moreover, most of today's conventional (dynamic) loudspeakers contain a permanent magnet, a metal coil behaving like an electromagnet, and a cone to translate an electrical signal into an audible sound [34], as shown in Fig. 2. When operating correctly, such a loudspeaker would naturally produce a magnetic field, originating from both the permanent magnet fixed inside the speaker, and the movable coil that creates a dynamic magnetic field when an electric current flows through it.

Therefore, our key insight is to detect the magnetic field produced by the conventional loudspeakers. By using the magnetometer (compass) in modern smartphones, we can distinguish between a human speaker and a computer loudspeaker, since the human vocal tract would not produce any magnetic field. As we show below, such observations will help us design and obtain a robust defense system with high accuracy. Moreover, we use the Spear speaker verification system as a building block to defend against the human impostor.

### *C. Use Cases*

To successfully leverage our key insight, we require users to place the smartphone as close as possible to the sound source. This is because the magnetic field produced by the loudspeaker can only be detected within a short range. However, the distance between the smartphone and the sound source is hard to measure. Therefore, we design non-intrusive use cases to confine the moving pattern of the smartphone and assist in measuring the distance. As shown in Fig. 3, our scheme requires the user first to open our mobile application and hold the smartphone near his head vertically or horizontally (a similar interaction model has been adopted by [11]); the user starts speaking the voice command while moving the smartphone towards his or her mouth at the same time. Finally, the user waits for our application to verify his identity. During this process, our application first collects the acoustic data and the reading of the inertial sensors, and then feeds them into the verification pipeline.

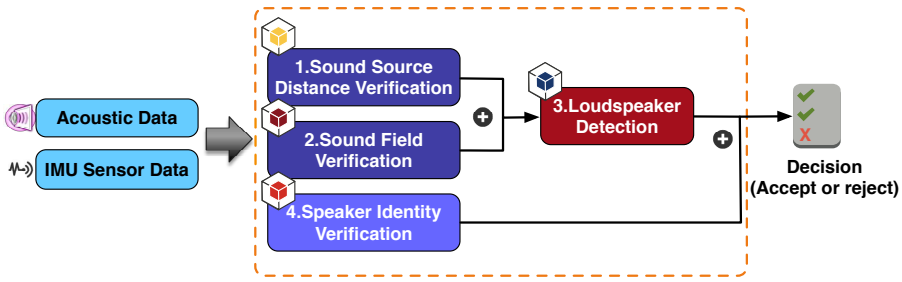



Fig. 4: The architecture of our defense system.

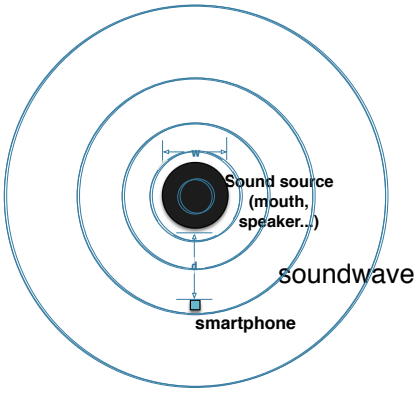



Fig. 5: Geometric constraint of our system

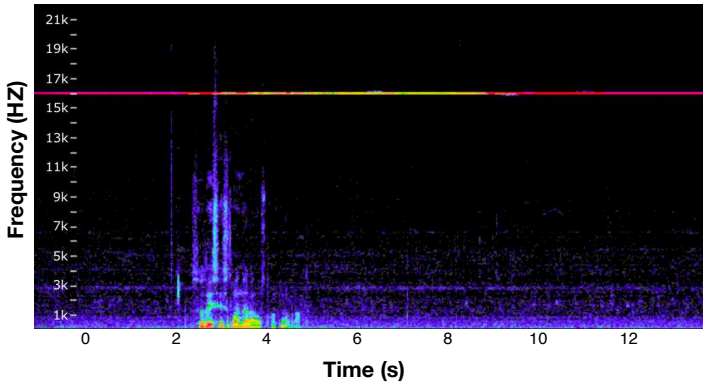



Fig. 6: Received spectrograph of the high-frequency tone while moving the phone.

## IV. The Proposed Solution

### A. System Architecture

As shown in Fig. 4, our system consists of four verification components for defending against voice impersonation attacks: 1) sound source distance verification, 2) sound field verification, 3) loudspeaker detection, and 4) speaker identity verification components.

The sound source distance verification component is designed for calculating the distance between the smartphone and the sound source. It manipulates the smartphone trajectory recovery algorithm with acoustic and sensory data to reconstruct the moving trajectory of the smartphone. We utilize the least-square circle fitting algorithm [17] to calculate the distance. The purpose of this component is to ensure that the smartphone is placed close enough to the sound source so that we can detect the magnetic field created by the loudspeaker with the smartphone built-in magnetometer.

The sound field verification component is designed for analyzing the characteristic of the sound field produced by the sound source. We add this element because the magnetometer is not sensitive enough to detect magnet in a small size, such as the magnet inside an earphone. Therefore, we use this component to detect if the sound is formed and articulated by a sound source, whose size is close to a human mouth (i.e., not a loudspeaker).

If the collected dataset passes the second and third tests, we then use the loudspeaker detection component to perform further detection. By cross-checking the magnetometer and motion trajectory data, we are able to verify if the sound is produced by a human speaker or a loudspeaker. The fourth component is designed for speaker identity verification, and is based on analyzing the spectral and prosodic features of the acoustic data. We leverage the state-of-the-art speaker verification algorithm to detect human-based voice impersonation attacks. Thus, combining the detection result from the fourth component with the one from the third component, we are able to defend against both machine-based voice impersonation attacks and human-based voice impersonation attacks on smartphones.

### B. Defending Against Machine-Based Voice Impersonation

*1) Sound Source Distance Verification:* As shown in Fig. 5, to calculate the distance $d$ between the sound source and the smartphone, we use speakers, microphones and inertial sensors to reconstruct the moving trajectory of the smartphone.

**Motion Trajectory Reconstruction.** As we mentioned before, we require the user to hold and move the smartphone toward his mouth while speaking. In the meantime, we collect both the acoustic data and the inertial sensor data from the smartphone. In our system, we adopt a similar phase-based distance measurement method as in [49] to calculate the distance using the following steps.

First, we let the smartphone's speaker generate inaudible tone in a static high frequency $f_s$ ( $f_s > 16$ kHz). Since the corresponding wavelength of that sound is less than 3 centimeter, the movement of the smartphone will significantly change

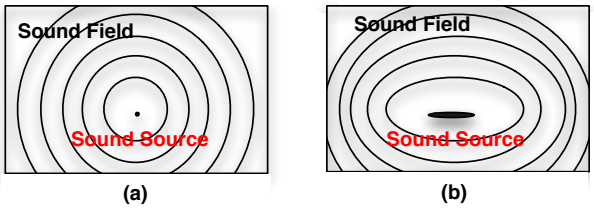



Fig. 7: The sound field created by (a) a point sound source and (b) created by a strip-type sound source.

the phase when it reflects off from the user's head. Based on the limitation of the speaker on commodity smartphones, we select the highest possible frequency using a calibration method described in [18]. With the high-frequency tone being broadcasted, the movement of the smartphone will cause phase change. Fig. 6 shows the received spectrograph of the high-frequency tone while moving the phone. Since the phase change is directly related to the moving distance $d$ of the smartphone, we can easily reconstruct the estimated moving trajectory and correlate it with the value derived from the inertial sensor.

Instead of tracking the smartphone in 3D space with free movement, we set up a pre-defined 2D moving plane. We assume the smartphone stays in the same plane while moving. The moving trajectory of the smartphone is approximate to a straight line, where the smartphone screen always faces the human's head while moving. Based on this model, we can use the time interval between the smartphone direction change combined with the relative moving speed to estimate the relative location of the smartphone in a 2D plane. As the magnetometer reading can result in some error in an indoor environment [37], we jointly use the magnetometer, gyroscope, and accelerometer to obtain the direction change $\Delta\omega$ [31].

By using the pre-defined 2D trajectory model, we can then set the start location as $(0, 0)$ and keep updating the location coordinate $(x_t, y_t)$ by combining the timestamp $t$, velocity $v$ and direction $\omega$ information. Finally, we can fully reconstruct the phone's 2D moving trajectory.

*2) Sound Field Verification:* In our defense system, we simplify the human voice as an acoustic sound source. Therefore, the user's speech is regarded as an acoustic signal broadcast by the sound source. The amplitude of the acoustic signal, which is the sound intensity level, can be measured by smartphone's microphone. To justify whether the received sound is broadcast from a human mouth, our system first models the *sound field* of the human mouth using the training data. Then, by performing a binary classification of each set of newly received sound data, we can verify the result. Therefore, only the sound source (or sound channel) with a similar size of a human mouth can be accepted and will be further processed.

**Quantifying the Sound Field.** The sound field represents the energy transfer in the air by the acoustic waves. The sound intensity level can express the energy contained in sound fields.

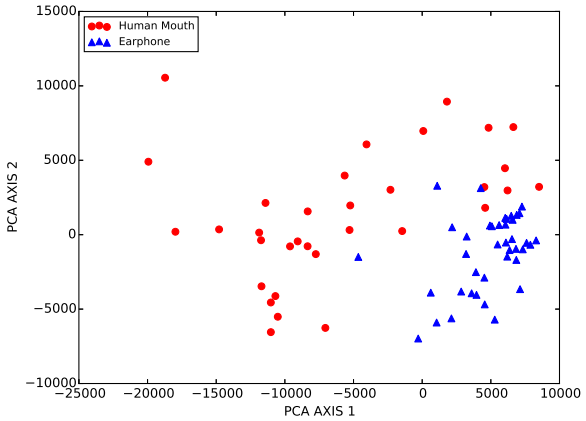



Fig. 8: The feature points of the human-mouth sound field (red circles) and the earphone sound field (blue triangles) after principal component analysis (PCA).

Fig. 7-(a)(b) shows the sound field created by a point sound source, and the sound field generated by a strip-type sound source, respectively. According to [19], the sound field around the user is affected not only by the vocal tract but also by the shape of the user's mouth and head. By allowing users to hold and horizontally move the phone in front of the sound source, we can collect a set of sound intensity measurements from different locations, which are further utilized to quantify the spatial characteristics of the sound field.

**Two Phases in Sound Field Verification.** As shown in Fig. 9, the sound source verification process is divided into two phases, the training phase and the predicting phase. In the training phase, we collect several sets of sound intensity as training data and use them to model the spatial characteristics of the user's sound field. While moving the smartphone as instructed, the user needs to speak the command displayed on the smartphone's screen repeatedly. For each round, we build a feature vector to represent the quantified sound field. Each feature vector contains multiple datasets, and each dataset is composed by a tuple of volumes (dB) and the rotation angle (degree). Specifically, the volume of the sound is measured by the microphone, and the rotation angle is jointly measured by the magnetometer, the gyroscope, and the accelerometer [37]. These feature vectors are then used to train a binary classifier using the linear Support Vector Machine (SVM) [12] algorithm. In the prediction phase, we ask users to perform a similar motion trajectory with the smartphone (as they did in the training phase). We then submit the newly collected feature vector to the pre-trained binary classifier and validate the results. Fig. 8 shows the feature vector of the human mouth sound field and the earphone sound field after applying the Principal Component Analysis (PCA) [52]. This shows that the feature points are easy to be separated, and thus the sound source size can be correctly classified.

*3) Loudspeaker Detection:* The goal of the loudspeaker detection component is to detect the emitted magnetic field. Unlike human vocal tract, conventional loudspeakers leverage

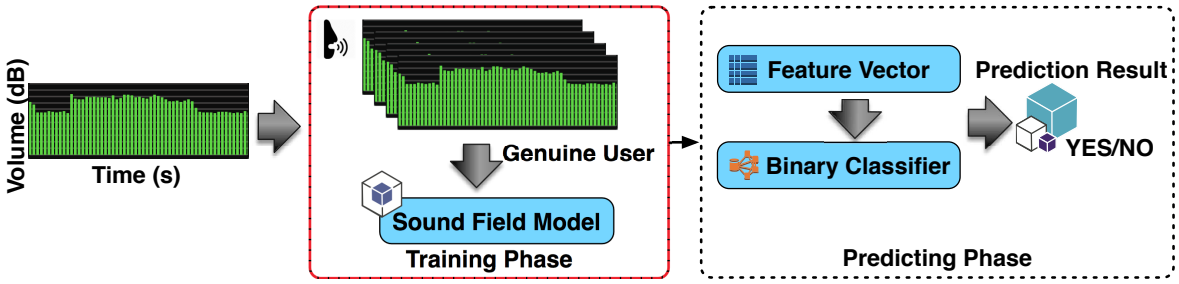



Fig. 9: The sound source validation process, containing two phases: i) Training phase and ii) Predicting phase.

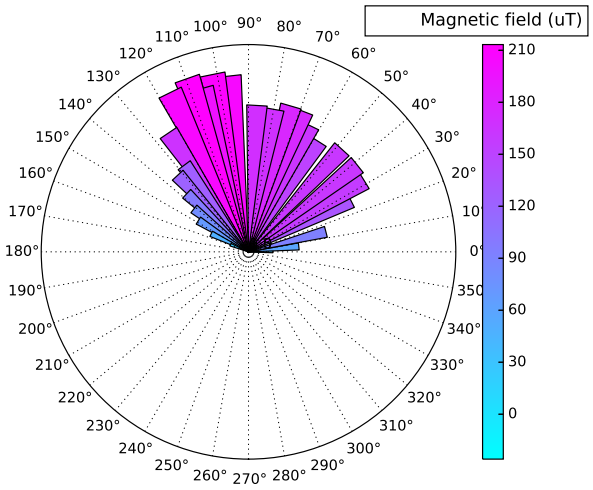



Fig. 10: The polar graph of the magnetic field reading for a conventional loudspeaker. (Note that the magnetic field strength emitted by loudspeakers usually ranges from 30 – 210$\mu T$).

magnetic force to transfer the electrical signal into acoustic sound. According to the validation mechanism presented above, two geometric constraints of the sound source and the smartphone in the submitted trajectory should be satisfied: i) the smartphone is close enough to the sound source, which means the distance is within a certain threshold $D_t$; ii) the size of the voice channel is close to the human mouth. Therefore, if an imposter tries to launch a machine-based impersonation attack using the loudspeaker, we can detect the speaker by checking the variance of the magnetometer readings.

Fig. 10 shows the polar graph (0°–180°) of the magnetic field reading for a conventional loudspeaker (Logitech LS21). Note that different loudspeaker may have different structure appearances as well as the magnetic field distributions. In our system, we jointly use the absolute value and the changing rate of magnetic readings to detect the speaker. We set a magnetic strength threshold $M_t$ and a changing rate threshold $\beta_t$. Both values are determined based on our experimental results.

### C. Defending Against Human-Based Voice Impersonation

*1) Speaker Identity Verification:* As part of our defense system, we choose the state-of-the-art Spear system as the speaker identity verification component to defend against human-based voice impersonation attacks. The Spear system has already implemented multiple mature speaker verification algorithms and has been evaluated using several standard voice datasets (e.g., Voxforge [48], NIST SRE [41] and MOBIO [28]). The toolchains provided by the Spear system are configurable. We further choose the *Gaussian Mixture Model* (GMM) and *Inter-Session Variability* (ISV) techniques. Spear has two phases, a training phase and a testing phase. Both phases require the voice data as an input. In our design, our application first collects the genuine user's voice samples to model the user using Spear (the voice samples are also used for the sound source verification), and then uses the trained speaker model to identify the incoming voice samples.

TABLE I: The performance of speaker identity verification component using the false acceptance rate (FAR).

| | Test 1 (FAR) | Test 2 (FAR) |
|---|---|---|
| **UBM** | 0.0% | 0.5% |
| **ISV** | 0.0% | 1.3% |

We evaluate the performance of the Spear system for defending against human-based voice impersonation attack by conducting two tests. For the first test, we create a dataset which consists of five speakers. Each speaker is asked to pronounce a unique six-digit passphrase for five times. We then allow the speaker to collect other speakers' voice samples and ask them to mimic it. Technically, the Spear system is for training and testing our data set. As shown in Table I, the false acceptance rates (FAR) for both of the GMM and ISV models are all equal to zero, which implies the success rate of the human-based voice impersonation attack is equal to zero. For the second test, we use the existing Voxforge dataset to train the Spear speaker model and test it using the CMU Arctic Database [22], in which they pronounce the same utterance when recording. The FAR value for the second test is significantly low, which confirms that Spear is very robust for defending against human-based voice impersonation attacks.

## V. Implementation

To evaluate and validate the effectiveness of our system, we build a prototype implemented on several smartphone testbeds from three different manufactures (shown in Table II), running Android 4.4 KitKat and one Arch Linux [6] server

TABLE II: Types of smartphones.

| Maker | Model |
|---|---|
| Google (LG) | Nexus 5 |
| | Nexus 4 |
| Samsung | Galaxy Nexus |

TABLE III: Four categories of output decisions.

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

with Intel(R) Core(TM) Devil's Canyon Quad-Core i7-4790K @ 4.00 GHz CPU and 32 GB of RAM.

Our prototype is based on a typical client-server architecture and can be divided into two parts: 1) a mobile application running on Android and 2) a server backend deployed in a virtual private cloud (VPC).

1) **Mobile Application**. The mobile application allows users to record and upload acoustic data annotated with inertial sensory information. We design and implement a simple graphical user interface (GUI) (Fig. 11) for guiding mobile users moving the smartphone while speaking the command.

2) **Server Backend**. The server backend has two main functionalities: i) handling incoming acoustic and inertial sensory data, and ii) processing received data and feeding back the verification decision. Our defense system uses a computer server configured with Arch Linux and Tornado web server [44] for parallel data processing.

**Handling Incoming Data.** We utilize a Tornado web server to process incoming connection requests. Tornado is a high-performance asynchronous web server, and it is capable of receiving and handling data from a larger number of users simultaneously. Our mobile clients send zipped data to the Tornado server via a secure web socket protocol and all the data sent from the users is encrypted to ensure confidentiality.

**Data Processing Pipeline.** At the server side, we first unzip the received data and then feed it into a cascade pipeline as we described in the previous section. Besides, we leverage the Advanced Python Scheduler (APScheduler) to accelerate the process of defending against the machine-based voice impersonation attack. The verification result is directly sent back to the smartphone through the secure web socket channel.

## VI. Evaluation

### A. Methodology

To perform our experiments, we design and build a small testbed environment with a real loudspeaker and a smartphone hardware. Because the Spear sub-system can address the human-based voice impersonation attacks, our evaluation focuses on the machine-based voice impersonation anti-spoofing sub-system. Since our method is for differentiating between a human speaker and a computer loudspeaker, we do not identify the differences among the voice replay attack, the voice morphing attack and voice synthesis attack as they all use the loudspeaker.

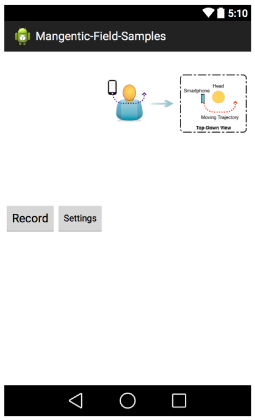
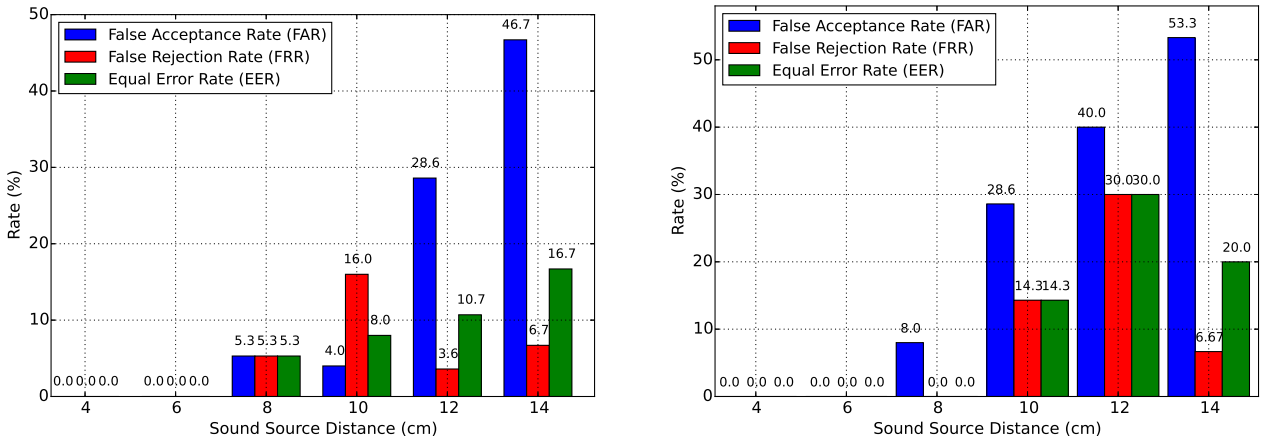



Fig. 11: The graphical user interface (GUI) for mobile user for guiding mobile users moving the smartphone while speaking the command.

**Devices and Tools.** We evaluate our system on smartphones. The models of smartphone testbed for implementing our system are shown in Table II. Appendix A provides the models of PC loudspeakers, notebook internal speakers, smartphone internal speakers, and earphones used in our evaluations.

**Performance Metrics.** As shown in Table III, our system contains four possible outcomes, where two are correct and two are incorrect. To assess the performance of our defense scheme, we choose the standard automatic speaker verification metrics, namely, the false acceptance rate (FAR) and the false rejection rate (FRR). FAR characterizes the rate at which an attacker is wrongly accepted by the system and considered as an authorized user. On the other hand, FRR characterizes the rate at which a true user is falsely rejected by our systems. Both FAR and FRR are controlled by adjusting the verification threshold. An attacker can launch a successful attack when the system confuses a spoofing attempt with a genuine one. In addition to FAR and FRR, we also measure the equal error rate (EER), which is the rate at which the acceptance and rejection errors are identical. To measure the EER for each test round, we vary the threshold value of each verification component in the defense scheme. A system with a perfect accuracy should have a zero EER.

**Sound Source Distance.** To assess the impact of the sound source distance in the defense mechanism, we create a test database which consists of five individual speakers. Each speaker contributes six groups of voice samples measured at different distances. We further use the recorded voice samples to perform machine-based voice replay attack using 25 different loudspeakers at various distances. The results coming from each of our system components are measured and merged. As shown in Fig. 12 (a), the FAR, FRR, and EER are all zero when the sound source distance is less than or equal to 6 cm. This is mainly because when the smartphone is placed very close to the loudspeaker, the magnetic field of the loudspeaker heavily interferes with the magnetometer's reading. Therefore, we can easily set up a threshold to differentiate the individual




(a) No shielding

(b) Magnetic field shielding.

Fig. 12: Impact of sound source distance for (a) No shielding and (b) Magnetic field shielding 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.

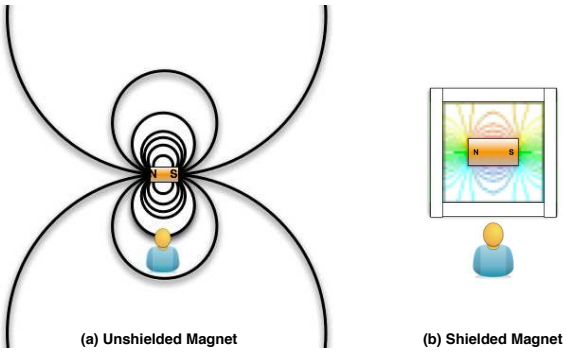


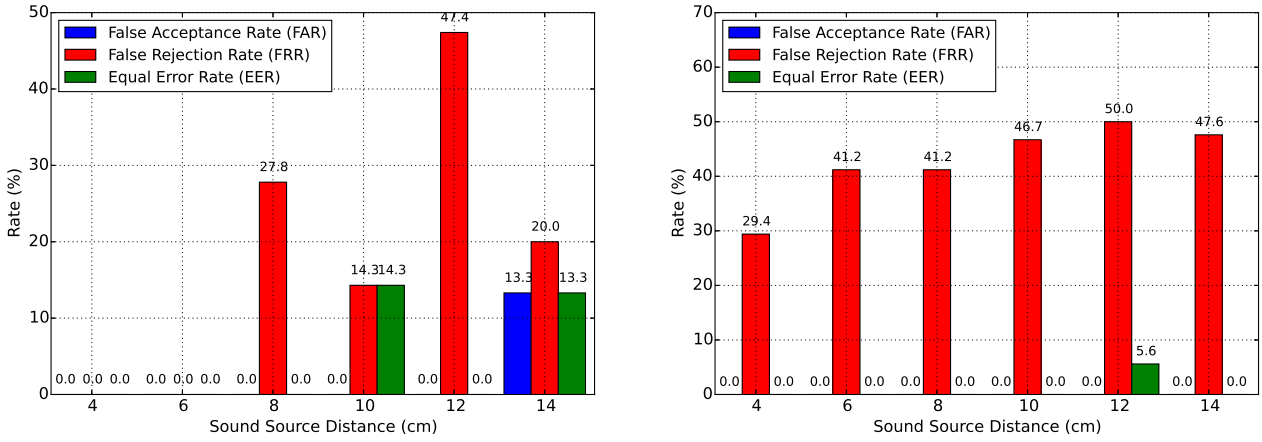

Fig. 13: The magnetic field distribution of: (a) unshielded magnet and (b) shielded magnet.

speaker and the loudspeaker. From 8 to 10 cm, the magnetic field emitted from the loudspeaker becomes weaker, and the FAR rises from zero to approximately 5%. When the distance between the smartphone and the sound source is larger than 10 cm, the magnetic field emitted from the loudspeaker becomes feeble, which is hard to differentiate from environmental magnetic interferences. Hence, the FAR rises sharply. However, the FRR remains low within all distance ranges (except at 10 cm) because the individual speaker does not produce the magnetic field. Thus, it can be correctly distinguished when there are no environmental magnetic interferences. According to the evaluation results, we set the sound source distance threshold $D_t$ to 6 cm for the best system performance.

**Magnetic Field Shielding.** Unlike the electrical field, the magnetic field can never be eliminated. One common way to avoid the emanation of the magnetic field is to use a metal (e.g. iron) box which covers the magnet. In this way, the magnetic field travels within the walls of the box and cannot penetrate the box (shown in Fig. 13). Among all the metals, the *Mu-metal* [1] achieves the best performance to shield the magnetic field. Mu-metal is a nickel-iron alloy, with 77% nickel, 16% iron, 5% copper, and 2% chromium. It has a high magnetic permeability that is perfect to shield the magnetic field.

To evaluate our system performance against machine-based voice impersonation attack using magnetic field shielding, the test database created from the sound source distance experiment is utilized. Different from the previous experiment, we now perform machine-based voice replay attack with the loudspeaker shielded by the Mu-metal. The results are measured from each of our system components and combined. As in Fig. 12 (b), the FAR, FRR, and EER values are equal to zero when the distance is less than or equal to 6 cm. This is because the metal box can still be detected by our system, as the magnetometer can detect both the magnet and the metal [45]. Moreover, the shielding metal also changes the sound field distribution of the loudspeaker, so our sound field validation component is still able to detect the anomaly. According to the results at 8 cm, the Mu-metal successfully decreases the magnetic field created by the loudspeaker and results in a higher FAR (8%) compared to the unshielded result (5.3%). From 8 to 14 cm, the values of FAR, FER, and EER increase dramatically as the Mu-metal significantly decreases the intensity of the magnetic field emanated from the loudspeaker. Based on these results, our system can be applied to detect shielded loudspeakers when the distance between the sound source and the smartphone is less than or equal to 6 cm.

**Environmental Magnetic Interference.** In order to assess the impact of environmental magnetic interference, we set up two test scenarios. First, the success rate of our method is evaluated when a user is nearby a computer. Same as in the previous experiments, we collect test data from both legitimate users and voice impostors with various distances. During the test, an all-in-one computer (iMac 27") is put 30 cm away from the test location. Hence, we expect high electromagnetic field (EMF) that may cause interference to our system. Before




(a) Near a computer

(b) In a car.

Fig. 14: 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).

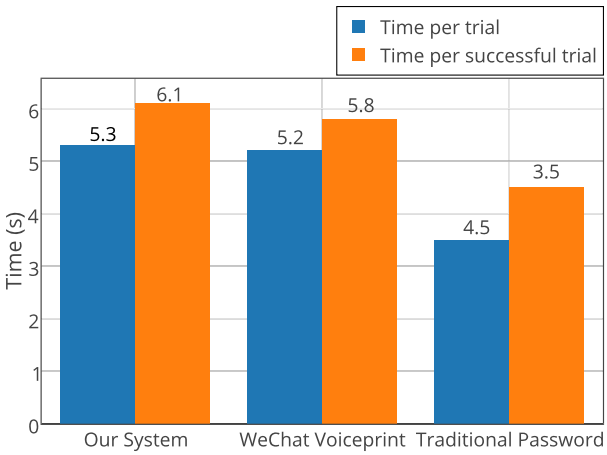



Fig. 15: Authentication time comparison.

conducting the experiment, we first measure the EMF radiation by using an Acoustimeter RF meter (Model AM-10) at the distance of 30 cm. The results show that the average exposure level varies from 500 $\mu W/m^2$ to 2500 $\mu W/m^2$. As shown in Fig. 14 (a), the FAR, FRR, and EER values are equal to zero when the distance is less than or equal to 6 cm. However, different from previous results, the FRR value rises sharply (27.8%) while the FAR remains at zero at the distance of 8 cm. This is mainly because, with the increase of the distance, the moving trajectories of the smartphone become closer to the computer screen, and the smartphone is exposed to heavier EMF radiation. Therefore, the interference from the EMF affects the reading of the magnetometer and triggers a false alarm.

Second, we conduct the same experiment in a car's front seat (Hyundai Sonata 2012). Since modern cars are equipped with many electronics, all of these electronics are emitters of EMF, potentially resulting in a very high level of EMF interference. As we expected, the evaluation result shown in Fig. 14 indicates that our method suffers a high FRR (around 45%) at a distance above 4 cm. Even at 4 cm, the FRR is still near 30%, which is unacceptable in our evaluation. However, the EERs in all test distances remain at zero. The results indicate that by adjusting the sensitivity level of the detection components (in particular, the loudspeaker detection component), we can achieve much better FAR and FRR results. Therefore, one solution could be by letting the smartphone sense the environment before collecting the data and adjusting its sensitivity level automatically. We will discuss more details of this solution later.

**Authentication Speed and Usability.** We compare the authentication time of our method, WeChat voice print, and credential based authentications. We recruit 20 volunteers (non-computer science background). Each of the volunteers performed ten trials of voice authentication using our system. In addition to the 200 trials in our system, our volunteers also performed 200 trials on WeChat voiceprint, as well as 200 trials to log in on WeChat using a traditional password. For all these experiments, we stop the time counter only when the authentication result is sent back. We try to minimize the influence of network latency by redirecting all network traffic to a local server and record the data transmission time. The time costs of the three schemes are averaged and plotted in Fig. 15 (Note that "Time per trial" contains unsuccessful trials which can be considered as false negatives). This figure indicates well that our system is only less than a second slower than the original WeChat voice print method. Moreover, both approaches are comparable to the traditional credential-based method.

**Various Classes of Speakers.** To demonstrate our proposed defense system is universal, we have selected and tested 25 different conventional loudspeakers ranging from low-end to

Fig. 16: Plastic CAB tube for sound-tube attack.

high-end, including PC loudspeakers, mobile phone internal speakers, laptop internal speakers, and earphones. For the lack of space, we omit the make and model information of those evaluated speakers and the detailed evaluation results. However, in short, the main result shows that our method can detect all of these loudspeakers owing to the same structure they share, all containing a permanent magnet. Thereby, the detection method should be the same. Besides, the magnetometer sensor AK8975 used by the smartphone has a sensitivity of $0.3\mu T/LSB$ and a measurement range of $\pm 1200\mu T$. On the other hand, as shown in Fig. 10, the magnetic field strength emitted by the loudspeakers is usually within the range of $30 - 210\mu T$. Therefore, the magnetic field based detection mechanism is quite reliable within a short distance.

## VII. Discussion

**Unconventional Loudspeakers.** Different from conventional loudspeakers which use magnetic force to create sound, some of the unconventional loudspeakers use an alternative way to produce a sound wave. These loudspeakers are usually very costly, and therefore unlikely to be adopted by a large population. However, as a defense system, we need to consider all possible attack vectors. We take the Electrostatic Loudspeaker (ESL) as an example of unconventional loudspeakers which does not produce a magnetic field. An electrostatic loudspeaker (ESL) consists of two metal grids with a plastic diaphragm. The diaphragm constantly charges a fixed positive voltage and creates a strong electrostatic field around it. It generates sound by the metal grids which are electrodes. Without utilizing the electrodynamic method to create sound, this type of speaker does not create a magnetic field. However, this kind of speaker can still be detected by magnetometer as the metal grids generate the magnetic interference. We notice that this type of loudspeakers usually has a larger size, which can also be detected by the sound field verification component. Another example is the Piezoelectric speakers which the electric current in the piezo crystal generates a movement (piezo effect) which produces the sound. Although it is already used by some phones, such speakers typically do not have good audio quality at the current stage.

**Sound-tube Attacks.** We further test our system against the sound-tube attacks. In this experiment, we ask volunteers to use several different size plastic CAB tubes (shown in Fig. 16) as "sound tube" and a loudspeaker to launch the attack. The plastic tube keeps a sufficient distance between the loudspeaker and the phone, and also transmits sound to break our sound field verification mechanism. However, all their attempts failed, mainly because replicating a human sound field using a mechanical device is hard to achieve. Furthermore, the attacker needs to cancel out sound resonance effect in the tube and simulate the shape of the mouth, which requires very sophisticated structure design.

**Adaptive Thresholding.** All four verification components in our defense scheme leverage thresholding to validate the input. We manually set the thresholds to achieve the best possible performance (FAR, FRR, EER) in a normal usage scenario. However, for some particular usage scenarios where the user is exposed to a high electromagnetic field (EMF) radiation, e.g., near a computer or in a car, adaptive thresholding may produce better results. As a future work, we propose the following solution: i) when encountering high environmental EMF radiation, we ask users to calibrate the smartphone by monitoring the environment for a few seconds, and ii) we calculate the average environmental magnetic interference level and adjust the threshold for each verification component adaptively. However, the design of this function should be with caution as it is possible to trick the application by training it at a high EMF environment, and then using the loudspeaker in a low EMF environment.

**Dual Microphones.** Certain smartphones like Nexus 4 have two microphones, and one of them is usually used for noise cancellation. To further improve the usability of our system, in the future we plan to utilize the dual microphones to reduce the required moving distance. The main idea is to measure the sound level difference (SLD) feature between the two microphones of the device. We then use sound volumes information with the SLD feature to perform sound field verification. Because different types of smartphones offer different dual-microphone layouts, we also need to investigate the estimation method for automatically setting the sound field verification parameters.

## VIII. Conclusions

This paper presents a robust software-only voice impersonation defense system tailored for smartphones and is readily deployable on existing mobile platforms. Our solution leverages the fact that the loudspeaker used in the machine-based voice impersonation attack has special physical charactertistics, i.e., it generates a magnetic field. We exploit this insight by non-intrusively requiring the user to place the smartphone near the sound source for detection and use the magnetometer to differentiate the human speaker and the loudspeaker. The prototype of our defense scheme achieves a nearly perfect accuracy and zero equal error rates in detecting the machine-based voice impersonation attack on smartphones. The experiment results show that our solution is capable of defeating the vast majority of voice impersonation attacks. Furthermore, our system significantly raises the level of security for existing voice-based mobile applications.

## IX. Acknowledgement

We thank the helpful comments from the anonymous reviewers. The first author's work was mainly done when he was at University at Buffalo, SUNY. This work was supported in

part by US National Science Foundation under grants CNS-1421903, CNS-1643207, and Global Research Lab. Program of Korea National Research Foundation under grant NRF-2016K1A1A2912757.

## REFERENCES


[1] Mu-metal. http://en.wikipedia.org/wiki/Mu-metal.

[2] A. G. Adami, R. Mihaescu, D. A. Reynolds, and J. J. Godfrey. Modeling prosodic dynamics for speaker recognition. In *ICASSP*, 2003.

[3] F. Alegre, A. Amehraye, and N. Evans. Spoofing countermeasures to protect automatic speaker verification from voice conversion. In *ICASSP*, 2013.

[4] F. Alegre, R. Vipperla, N. Evans, and B. Fauve. On the vulnerability of automatic speaker recognition to spoofing attacks with artificial signals. In *EUSIPCO*, 2012.

[5] T. B. Amin, J. S. German, and P. Marziliano. Detecting voice disguise from speech variability: Analysis of three glottal and vocal tract measures. *The Journal of the Acoustical Society of America*, 134(5):4068–4068, 2013.

[6] ArchLinux. https://www.archlinux.org/.

[7] audioBoo. https://audioboom.com/.

[8] Baidu. Voice biometric on smartphone. http://shouji.baidu.com/.

[9] T. Bin Amin, P. Marziliano, and J. S. German. Glottal and vocal tract characteristics of voice impersonators. *IEEE Transactions on Multimedia*, 16(3):668–678, 2014.

[10] J. P. Campbell Jr. Speaker recognition: a tutorial. *Proceedings of the IEEE*, 85(9):1437–1462, 1997.

[11] S. Chen, A. Pande, and P. Mohapatra. Sensor-assisted facial recognition: an enhanced biometric authentication system for smartphones. In *ACM MobiSys*, 2014.

[12] N. Cristianini and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. 2000.

[13] K. Delac and M. Grgic. A survey of biometric recognition methods. In *Electronics in Marine*, 2004.

[14] W. Diao, X. Liu, Z. Zhou, and K. Zhang. Your voice assistant is mine: How to abuse speakers to steal information and control your phone. In *ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*, 2014.

[15] G. R. Doddington et al. Speaker recognition based on idiolectal differences between speakers. In *INTERSPEECH*, 2001.

[16] Festvox. http://festvox.org/.

[17] W. Gander, G. H. Golub, and R. Strebel. Least-squares fitting of circles and ellipses. 3(5):63–84, 1996.

[18] S. Gupta, D. Morris, S. Patel, and D. Tan. Soundwave: using the doppler effect to sense gestures. In *SIGCHI Conference on Human Factors in Computing Systems*, 2012.

[19] B. F. Katz and C. D'Alessandro. Measurement of 3d phoneme-specific radiation patterns in speech and singing. Scientific Report, 2007.

[20] C. S. Kei and Jack. Superlock. https://code.google.com/p/voiceprint-model-builder-for-superlock/downloads/list.

[21] E. Khoury, L. E. Shafey, and S. Marcel. Spear: An open source toolbox for speaker recognition based on bob. In *ICASSP*, 2014.

[22] J. Kominek and A. W. Black. The cmu arctic speech databases. In *Fifth ISCA Workshop on Speech Synthesis*, 2004.

[23] Kong Aik Lee, Bin Ma, and Haizhou Li. Speaker verification makes its debut in smartphone. *IEEE Signal Processing Society Speech and language Technical Committee Newsletter.*, 2013.

[24] K. B. Lee and R. A. Grice. The design and development of user interfaces for voice application in mobile devices. In *International Professional Communication Conference*, 2006.

[25] J. Lindberg, M. Blomberg, et al. Vulnerability in speaker verification-a study of technical impostor techniques. volume 99, pages 1211–1214, 1999.

[26] J. Mariéthoz and S. Bengio. Can a professional imitator fool a gmm-based speaker verification system? Technical report, 2005.

[27] H. Melin. Automatic speaker verification on site and by telephone: methods, applications and assessment. 2006.

[28] Mobio. https://www.idiap.ch/dataset/mobio.

[29] J. Y. Nicholas Evans and T. Kinnunen. Spoofing and countermeasures for speaker verification: a need for standard corpora, protocols and metrics. *IEEE Signal Processing Society Speech and language Technical Committee Newsletter.*, 2013.

[30] Nuance. Nuance vocalpassword. http://www.nuance.com/landing-pages/products/voicebiometrics/vocalpassword.asp, 2013.

[31] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen. Zee: zero-effort crowdsourcing for indoor localization. In *Mobicom*, 2012.

[32] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM systems Journal*, 40(3):614–634, 2001.

[33] R. Rawassizadeh, B. A. Price, and M. Petre. Wearables: has the age of smartwatches finally arrived? *Communications of the ACM*, 58(1):45–47, 2014.

[34] D. Reese, L. Gross, and B. Gross. *Audio Production Worktext: Concepts, Techniques, and Equipment*. 2012.

[35] D. Reynolds, W. Andrews, J. Campbell, J. Navratil, B. Peskin, A. Adami, Q. Jin, D. Klusacek, J. Abramson, R. Mihaescu, et al. The supersid project: Exploiting high-level information for high-accuracy speaker recognition. In *ICASSP*, 2003.

[36] J. Rodgers. Adobe voco - should we be afraid? http://www.pro-tools-expert.com/home-page/2016/11/16/adobe-voco-should-we-be-afraid.

[37] N. Roy, H. Wang, and R. Roy Choudhury. I am a smartphone and i can tell my user's walking direction. In *Mobisys*, 2014.

[38] W. Shang and M. Stevenson. Score normalization in playback attack detection. In *ICASSP*, 2010.

[39] M. Shirvanian and N. Saxena. Wiretapping via mimicry: Short voice imitation man-in-the-middle attacks on crypto phones. In *ACM CCS*, 2014.

[40] E. Shriberg, L. Ferrer, S. Kajarekar, A. Venkataraman, and A. Stolcke. Modeling prosodic feature sequences for speaker recognition. *Speech Communication*, 46(3):455–472, 2005.

[41] N. SRE. http://www.itl.nist.gov/iad/mig/tests/spk.

[42] Y. Stylianou. Voice transformation: a survey. In *ICASSP*, 2009.

[43] R. Togneri and D. Pullella. An overview of speaker identification: Accuracy and robustness issues. *Circuits and systems Magazine*, 11(2):23–61, 2011.

[44] Tornado. http://www.tornadoweb.org.

[45] D. Vandermeulen, C. Vercauteren, and M. Weyn. Indoor localization using a magnetic flux density map of a building. In *AMBIENT*, 2013.

[46] J. Villalba and E. Lleida. Detecting replay attacks from far-field recordings on speaker verification systems. In *Biometrics and ID Management*. 2011.

[47] J. Villalba and E. Lleida. Preventing replay attacks on speaker verification systems. In *ICCST*, 2011.

[48] Voxforge. http://www.voxforge.org/.

[49] W. Wang, A. X. Liu, and K. Sun. Device-free gesture tracking using acoustic signals. In *Mobicom'16*, 2016.

[50] Z.-F. Wang, G. Wei, and Q.-H. He. Channel pattern noise based playback attack detection algorithm for speaker recognition. In *ICMLC*, 2011.

[51] WeChat. Voiceprint. http://thenextweb.com/apps/2015/03/25/wechat-on-ios-now-lets-you-log-in-using-just-your-voice/.

[52] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1):37–52, 1987.

[53] Z. Wu, N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre, and H. Li. Spoofing and countermeasures for speaker verification: a survey. *Speech Communication*, 66:130–153, 2015.

[54] Z. Wu, T. Kinnunen, N. Evans, and J. Yamagishi. Asvspoof 2015: Automatic speaker verification spoofing and countermeasures challenge evaluation plan. *Training*, 10(15):3750, 2014.

[55] Z. Wu and H. Li. Voice conversion and spoofing attack on speaker verification systems. In *APSIPA*, 2013.

[56] H. L. Zhizheng Wu, Eng Siong Chng. Detecting converted speech and natural speech for anti-spoofing attack in speaker recognition. In *Interspeech*, 2012.

## APPENDIX A
## MODELS OF LOUDSPEAKERS USED FOR EVALUATION

TABLE IV: Makers and Models of Loudspeakers Used for Evaluation

| **Maker** | **Model** |
|---|---|
| Logitech 7 Watts RMS (FTC) 2.1 Stereo Speaker System | LS21 |
| Klipsch - 2-Way Indoor/Outdoor Speakers | KHO-7 |
| Insignia - 2-Way Indoor/Outdoor Speakers | NS-OS112 |
| Sony - Portable Bluetooth Speaker | SRSX2/BLK |
| Bose - SoundLink Mini Bluetooth Speaker | PINK |
| Bose - 151 SE(R) Environmental Speakers | 151 SE |
| Yamaha - Natural Sound 5” Outdoor Speakers | NS-AW190BL |
| Pioneer - 5-1/4” Floor Speaker | SP-FS52 |
| HP - 2.0 Speaker System | D9J19AT |
| GPX - 2.1 Speaker System | HT12B |
| Coby - 2.1 Home Audio Speaker System | CSMP67 |
| Acoustic Audio - AA2101 | AA2101 |
| Macbook Pro (Mid 2012) Internal Speaker | A1286 |
| Macbook Air (Mid 2012) Internal Speaker | A1466 |
| iMac (Late 2009) Internal Speaker | MB952XX/A |
| HP 6510b Internal Speaker | GM949 |
| Toshiba - Satellite Internal Speaker | C55-B5101 |
| Dell - Inspiron 5000 Series Internal Speaker | I5558-2571BLK |
| Apple iPhone 6 Plus Smartphone Internal Speaker | A1524 |
| Apple iPhone 5S Smartphone Internal Speaker | A1533 |
| Apple iPhone 4S Smartphone Internal Speaker | A1387 |
| LG Nexus 5 Smartphone Internal Speaker | LG-D820 |
| LG Nexus 4 Smartphone Internal Speaker | LG-E960 |
| Samsung Galaxy S Headset Earphones | EHS44 |
| Apple White 3.5mm Connector EarPods | MD827LL/A |