Speechless: Analyzing the Threat to Speech Privacy from Smartphone Motion Sensors

S Abhishek Anand
Department of Computer Science
University of Alabama at Birmingham
Email: anandab@uab.edu

Nitesh Saxena
Department of Computer Science
University of Alabama at Birmingham
Email: saxena@uab.edu

***Abstract*—According to recent research, motion sensors available on current smartphone platforms may be sensitive to speech signals. From a security and privacy perspective, this raises a serious concern regarding sensitive speech reconstruction, and speaker or gender identification by a malicious application having *unrestricted* access to motion sensor readings, *without* using the microphone.**

In this paper, we revisit this important line of research and closely inspect the effect of speech on smartphone motion sensors, in particular, gyroscope and accelerometer. *First*, we revisit the previously studied scenario (Michalevsky et al.; USENIX Security 2014), where the smartphone *shares* a common surface with a loudspeaker (with subwoofer) generating speech signals. We observe some effect on the motion sensor signals, which may indeed allow speaker and gender recognition to an extent. However, we also argue that the recorded effect on the sensor readings is possibly from *conductive vibrations* through the shared surface instead of direct acoustic vibrations due to speech as perceived in previous work. *Second*, we further extend the previous work by analyzing the effect of speech produced by (1) *other less powerful speakers* like the in-built laptop and smartphone speakers, and (2) *live humans*. Our experiments show that in-built laptop speakers were only able to affect the accelerometer when the laptop and the motion sensor shared a surface. Smartphone speakers were not found to be powerful enough to invoke a response in the motion sensors through aerial vibrations. We also report that in the presence of live human speech, we did not notice any effect on the motion sensor readings.

Our results have two-fold implications. First, human-rendered speech seems potentially incapacitated to trigger smartphone motion sensors within the limited sampling rates imposed by the smartphone operating systems. Second, it seems that even machine-rendered speech may not be powerful enough to affect smartphone motion sensors through the aerial medium, although it may induce vibrations through a conductive surface that these sensors, especially accelerometer, could pick up if a relatively powerful speaker is used. Overall, our results suggest that smartphone motion sensors may pose a threat to speech privacy only in some limited scenarios.

***Keywords*-side-channel attacks; motion sensors; speech privacy**

I. Introduction

Recent developments in the mobile device industry have seen an increase in the capabilities of the smartphone hardware to support applications that provide a comprehensive user experience. Motion sensors have played an important role in this task by collecting information about a user's activity, movement, and orientation. *Accelerometers* and *gyroscopes* are two of the most commonly used sensors on these devices that measure the motion and orientation of the device. However, recent studies have suggested a security flaw in these sensors by noticing a sensitiveness towards low frequency audio signals (specifically, speech). In particular, it has been believed that there exists a possibility of turning these sensitive sensors into microphones for picking up speech signals [1], [2].[1]

The possibility of turning motion sensors into microphones, capable of recording speech, has very adverse real-world implications. These motion sensors are readily available on smartphones and other smart wearable devices that have become a predominant feature in everyone's life. A unique fact about motion sensors on current smartphone platforms, specifically Android, is their unrestricted access. An application does not require special permission from the user to access the motion sensor readings. Hence, a malicious application, by obtaining access to motion sensor readings, may be able to achieve a similar threat level as directly accessing the microphone (that requires explicit permission) and recording an unsuspecting victim's voice or conversation.

The security threats of a malicious application gaining access to an unsuspecting victim's voice or conversation are particularly devastating. Sensitive information can be leaked in surreptitious manner if the malicious application is able to reconstruct speech from motion sensor readings. For example, sensitive verbal communications would be exposed, including information such as credit card numbers and social security numbers as the victim speaks into or near the phone such as over a phone call. In addition, various aspects of the eavesdropped speech signals can be utilized for speaker and gender identification. This threat violates the privacy of the victim(s) by revealing the identity and gender information that may otherwise be considered personal and should not be revealed unless proper permission has been granted by the involved parties.

In addition to *human-rendered speech*, *machine-rendered speech* also has the potential to be exploited by gaining access to motion sensors. A particular example could be closed auditorium or a meeting hall that is soundproof in order to avoid acoustic eavesdropping from outside world. If speech

[1]The work reported in [2] proposes a benign use case of detecting "hot keywords" for voice commands based on accelerometer. In this paper, our focus is on malicious use case of detecting speech through both gyroscope [1] and accelerometer readings.

DOI 10.1109/SP.2018.00004

is being communicated inside the hall using loudspeakers, a smartphone placed near the loudspeaker may pick up speech from the loudspeakers through its motion sensors. Thus, the attacker may be able to compromise the speech privacy of those present in the meeting. Since loudspeakers serve to amplify the sound, the effect on motion sensors may be more pronounced than live human speech thereby facilitating the attacker's task. Other examples for such a scenario could be private speeches or dinners where speech privacy is an essential requirement.

Another scenario of *machine-rendered speech* leakage may involve a user placing their smartphone in vicinity of their laptop while using the laptop's speakers. In this scenario, the sound from the laptop's speakers could possibly be picked up by the motion sensors of the user's smartphone. An extension to this case would be speech emanating from the smartphone speakers that could be picked up by the motion sensors of a smartphone. The in-built laptop and smartphone speakers are usually less powerful (than loudspeakers) which may have an impact on the motion sensors' capability to record speech generated by such commodity speakers.

In this work, we systematically explore the reaction of motion sensors to speech signals under different environments. We specifically consider possible threat scenarios and mark out the ones where speech privacy attacks that exploit motion sensors may be possible. As such, we do not build or improve upon any such attacks that already exist in literature. Our assumption is that in all our potential threat scenarios, exploiting motion sensors for speech recognitions (as proposed in [1], [2]) would be feasible and their accuracies can potentially be improved with sophisticated machine learning, increased sampling rate, and an increased sample size. In the initial phase, we revisit representative prior scenarios that were investigated in [1], [2] using *machine-rendered speech*. Since *machine-rendered speech* can be adjusted to have varying levels of loudness, these scenarios serve to outline the requirements that should be met in order to have a noticeable effect on the motion sensors. In addition, we design and study the effect of speech on motion sensors in other novel settings that have not been considered before, to explore the broader reach of this category of threat to user's speech privacy.

Further, we consider a natural scenario of speech signals rendered by a human subject in close proximity of the phone. Since the assumptions of previous work [1], [2] suggest that the motion sensors are responsive to low frequency audio signals (especially speech) to some extent, we measure the susceptibility of the motion sensors against human speech. The scenario involving live human speech has not been reported in prior research [1], [2].[2]

Our Contributions: We dissect the threat to speech privacy using motion sensors and assess its realism by analyzing the scenarios in which speech signals, traveling through air or conductive solid surfaces, affect the motion sensor readings. We do not seek to extend or improve upon the attacks developed in [1], [2] rather find out the scenarios where such attacks may be deployed successfully for compromising speech privacy. We believe that our work makes the following key contributions:

1) ***Effect of Machine-Rendered Speech on Smartphone Motion Sensors***: We observe the effect of machine rendering of the human speech on accelerometer and gyroscope readings in several settings. Using frequency and/or time domain analysis, we show that while there was no apparent change in sensor readings when there was no interaction between the physical world of the machine and the smartphone, some effect was observed when the machine and the device shared a surface (this setting mimics the scenario tested in [1][3]).
   This analysis seems to indicate that the sensors picked up *surface vibrations or conductive vibrations*, but *not* the acoustic aerial vibrations. As accelerometer sensor has been shown to pick up surface vibrations of keystrokes, making it possible to decode the typed information [3], we suggest that the surface, on which the smartphone and the speech rendering device is placed, plays an important role in the conduction of sound alongside the capability of the speech generating device. In particular, we find that relatively powerful loudspeakers (such as subwoofers) may be able to create such noticeable vibrational effects. In addition, in-built laptop speakers were able to produce faint response in the accelerometer when the laptop and the motion sensor shared a surface. Smartphone speakers were not powerful enough to invoke a response in the motion sensors through aerial vibrations. This analysis is presented in Section VI-B-VI-D.

2) ***Effect of Live Human Speech on Smartphone Motion Sensors***: We measure the effect of human speech on the motion sensor readings (Section VI). Using same methodology as above, we did not notice any significant changes in the motion sensors' measurements indicating that these sensors may not be significantly impacted by the human-rendered speech signals. We validate this result with a number of male and female speakers. This analysis is presented in Section VI-E.

Implications and Significance of Our Work: We believe that our work has important implications (a summary of our primary insights is captured in Section VII Table I). Our first key result is that *human-rendered speech* may potentially be incapable of triggering the smartphone motion sensors within the limited sampling rates (200Hz for Android) imposed by the mobile operating systems. This fact may bear good news for the security community since zero-permission motion sensors may not be exploited for directly deducing sensitive live speech spoken by a human entity.

Our second major insight is that even *machine-rendered*

[2]Confirmed via personal communication with 1st author of [1], [2].

[3]Confirmed via personal communication with 1st author of [1].

speech may not be powerful enough to impact smartphone motion sensors "through air", although it may induce vibrations through a conductive surface that these sensors could pick up. However, we believe that *conducting vibrations* represents an indirect (possibly less common) threat scenario involving a relatively powerful speaker that may also be a positive implication to the field of speech privacy.

Overall, the broader significance of our work is deconstructing the perception in the community that motion sensors can be exploited to compromise *human conversations* on smartphones. *Live human speech* was not tested in [1] while we have included different scenarios including human speech and speech through laptops and smartphone speakers. We show that the threat perceived by [1] does not go beyond loudspeaker/Laptop-Same-Surface scenarios. Such perceived threats raised by the potential of speech construction, speaker identification, and gender identification based on zero-permission motion sensors would have serious implications to the society as a whole. However, given that the research pointing to these threats is in its nascent stage [1], [2], it is very important to examine in detail the threat posed by this type of side channel attack under common use case scenarios where motion sensors could be exploited thereby threatening speech privacy.

Specific to the context of speech inference through motion sensors, it is important to re-validate the threat, especially given that the notion of this threat is appealing to people and has already made a significant impact through media coverage in premium outlets worldwide [4], [5], [6], [7], [8], [9], [10], [11]. This may have created a sense of insecurity among the readers. Our work, on the contrary, shows that speech inference, speaker identification and gender identification based on current smartphone motion sensors may not be feasible in all situations, given that *human-rendered speech* does not seem to have a direct effect on the readings of these sensors in such conditions. The *machine-rendered speech* effect seems limited to conductive vibrations, which are dependent on the contact surface and the audio source.

II. Background and Related Work

Motion sensors are a small piece of technology that measure and record a physical, motion-relevant property. This measurement or reading is then utilized by an application for required purposes. Accelerometers and gyroscopes are the common motion sensors deployed on smartphones. An accelerometer is used to measure movement and orientation and the gyroscope is used to measure angular rotation, across x, y, and z axes.

Motion sensors have been shown prone to acoustic noise particularly at high frequency and power level in [12], [13], [14], which showed that MEMS gyroscopes are susceptible to high power, high frequency noise that contains frequency components in proximity of the resonating frequency of the gyroscope's proof mass. This concept of work was further utilized by Son et al. [15] to interfere with the flight control system of a drone using intentional sounds that were produced by a Bluetooth speaker attached to the drones with a sound pressure level of 113dB. This attack was enough to destabilize one of the target drones used in the experiment due to fluctuations in the output of the gyroscope from the interference of the noise near the resonant frequency of the sensor.

The use of motion sensors (gyroscope, in particular) as a microphone to pick up speech signals was first reported by Michalevsky et al. [1]. They showed that the gyroscope sensor in smartphones might be sensitive enough to be affected by speech signals. Since gyroscope sensor in smartphones has a sampling rate of 200Hz, there exists an overlap with the frequency range of human voice especially at the lower end of the spectrum.

In another work done by Zhang et al. [2], it was shown that accelerometer readings could be affected by speech. In particular, they reported that it was possible to detect the voice commands (hotwords) spoken by the user through the accelerometer sensor.

Both [1] and [2] used speech that was produced by either a loudspeaker or a phone speaker to test its effect on the sensors. The Gyrophone [1] setup tested the impact of speech generated by a loudspeaker (with a subwoofer) on a phone placed on the same surface as the loudspeaker. AccelWord [2] tested the impact of speech generated by the phone speaker. We re-investigate both approaches in our work and extend them to other possible scenarios that have not been studied before.

In addition, there are examples of motion sensors leaking information other than speech, thereby compromising user privacy through another class of attacks. Cai et al. [16] used motion sensors to infer keystrokes from virtual keyboards on smartphone's touchscreen. Using vibration patterns from different parts of the keyboard, they were able to recover more than 70% of the keystrokes. This work was extended by Owusu et al. [17] by extracting 6-character passwords by logging accelerometer readings during password entry. Xu et al. [18] performed a similar study and were able to extract confidential user input (passwords, phone numbers, credit card details etc.) using motion sensors. In a work similar to [19], Miluzzo et al. [20] showed that it was possible to identify tap location on smartphone's screen with an accuracy of 90% and english letters could be inferred with an accuracy of 80%.

III. Motion Sensor Design

Motion sensors in smartphone and other smart devices are implemented as micro-electro-mechanical system (MEMS) that uses miniaturized mechanical (levers, springs, vibrating structures, etc.) and electro-mechanical (resistors, capacitors, inductors, etc.) elements developed using microfabrication. They are designed to work in coordination to sense and measure the physical properties of their surrounding environment.

MEMS Gyroscope: A gyroscope is a motion-sensing device, based on the principle of conservation of momentum, that can be used to measure angular velocity. An MEMS gyroscope works on the principle of rotation of vibrating objects or Coriolis effect [21]. This effect causes a deflection to the path of the rotating mass when observed in its rotating reference frame. MEMS gyroscopes fall in the category of

vibrating structure gyroscope as they use a vibrating mass in their design. The Coriolis effect described above causes the vibrating mass to exert a force that is read from a capacitive sensing structure supporting the vibrating mass.

MEMS Accelerometer: An accelerometer is an electro-mechanical device that can be used to measure gravity and dynamic acceleration such as motion and vibrations. The basic design of an MEMS accelerometer can be modeled as mass-spring system. A proof mass (an object of known quantity of mass) is attached to a spring of known spring constant, which in turn is attached to the support structure. An external acceleration causes the proof mass to move, causing a capacitive change that is measured to provide the acceleration value. It may also be noted that the accelerometer does not measure the rate of change of velocity rather it measures acceleration relative to gravity or free-fall.

IV. Preliminaries and Attack Scenarios

In this section, we discuss some preliminary notions that will be used in our analysis of motion sensor behavior in the presence of speech. We also examine the signal characteristics of speech and the response of motion sensors in the frequency range of the speech. We further look at scenarios that could be potential avenues for executing a side channel attack against speech privacy by exploiting the motion sensors.

A. Basic Audio Principles

The fundamental frequency for speech is between 100Hz to 400Hz. The fundamental frequency for a human male speech lies in the range 85Hz-180Hz and for a human female from 165Hz-255Hz. The fundamental frequency may change while singing where it may range from 60Hz to 1500Hz [22]. The sampling frequency of the MEMS sensors could range up to 8kHz. For example, the sampling frequency (also referred to as output data rate) in the latest Invensense motion sensor chip MPU9250 is described as 8kHz for the gyroscope and 4kHz for the accelerometer [23]. However, the operating platforms on smartphones often place *a limit on the sampling frequency* of these devices. This limit is often implemented in the device driver code and is *200Hz* for Android platform [1], [2] in order to prevent battery drain from frequent updates.

Nyquist sampling theorem states that to capture all the information about the signal, sampling frequency should be at least twice the highest frequency contained in the signal. For the MEMS motion sensors embedded in the smartphones, the sampling frequency is restricted to 200Hz that implies that they can only capture frequencies up to 100Hz. Hence, the motion sensor may only be able to capture a small range of the human speech in the sub-100Hz frequency range although due to aliasing effect we can expect higher frequency speech to feature in the sub-100Hz range as reported in [1].

B. Experimental Attack Scenarios

In order to test the effect of speech on MEMS motion sensors, we conceptualize different scenarios that encompass the intended objective of this work. There are three factors that should be taken into account in the experiments that affect the behavior of the motion sensors: (1) Source of speech, (2) Medium through which the audio travels, and (3) Pressure level of speech.

1) ***Source of Speech:*** Speech can be generated through various sources that we broadly classify into two categories: human voices and machine-rendered speech. Human voices could further be broken down into male voices and female voices. Machine-rendered speech involves rendering of a human voice through a speaker system. In our experiments, we use (a) a powerful speech generating device like a conventional loudspeaker with subwoofers (that boost low frequency sounds and may induce vibrations), and (b) in-built laptop speakers and smartphone speakers that are less powerful, as possible sources of speech.
2) ***Audio Transfer Medium:*** To consider the effect of speech on motion sensors, we need to take in account the medium through which an audio signal travels to the motion sensors. The transmission of speech to motion sensors could be through vibrations in the air or vibrations within the surface shared by both the speech generating device and the motion sensors. We test conduction of sound through air and through commonly used surfaces such as wood and plastic.
3) ***Sound Pressure Level:*** Sound pressure level is an indicator of the loudness of sound and is measured in decibels (db). Louder sounds contain more energy and could have greater effect on the motion sensors. For this reason, we test the sounds at different loudness measured in decibels to correlate loudness with effect on the motion sensors.

We design our scenarios based on the three factors detailed above. The initial setup in our work is similar to the experimental setup designed in [1] where the smartphone is placed on a desk with a loudspeaker that emits speech. For human speech, we position a human speaker very close to the desk on which the smartphone is placed to test the potential for capturing human speech.

1) Machine-Rendered Speech: We begin by recreating the scenario reported in Gyrophone [1], where the smartphone is placed on a desk with a loudspeaker (with subwoofer) that emits human speech. The scenario, henceforth referred as *"Loudspeaker-Same-Surface"*, is depicted in Appendix Fig 1. Here, the phone is in full contact with the surface on which the loudspeaker is placed. As motivated earlier, this scenario can occur in restricted closed door meetings or speeches where the designated speakers are speaking in a microphone and their speech is relayed to the audience through loudspeakers. In this case, the attacker places a smartphone on the same surface as the loudspeaker so that the motion sensors in the smartphone can pick up speech played through the loudspeakers, which are then read by the attacker. The attacker can also utilize a compromised smartphone that the user inadvertently places on the same surface as the sound source.

An additional scenario for machine-rendered speech would

be placing the smartphone containing the motion sensors on a different surface than the speech rendering device. We implement this scenario, called *"Loudspeaker-Different-Surface"*, by placing the smartphone on a different surface than the loudspeaker, as depicted in Appendix Figure 2. Additional scenarios are tested with laptop speakers *"Laptop-Same-Surface"*. Laptop speaker scenario can occur when the victim is in, for instance, a VoIP call using his/her laptop with its speakers turned on and put down their smartphone near the laptop. We also test smartphone speaker scenario *"Phone-Different-Surface"* similar to [2] where the speech is rendered through smartphone speakers and picked up by another smartphone placed in its vicinity.

2) Human Speech: In all the previously described scenarios, the speech used for measuring the response of the motion sensors is being produced by a loudspeaker. Such machine-rendered speech is different from a human speaker in the sense that a loudspeaker can effectively produce a louder speech than a human can. In order to achieve commonly occurring setup, we design a human speaker scenario where a human speaker speaks directly in the smartphone. This setup mimics a scenario where an attacker may eavesdrop on user's conversation that takes place on or near their smartphone.

In our experiment, we place the phone on a stationary and isolated surface and ask the test subjects to speak into the smartphone. In one scenario, we ask the human subjects to speak in normal voice (*"Human-Normal"*) and in the other scenario, we ask them to speak as loud as possible (*"Human-Loud"*) to maximize the effect of speech (if any) on the motion sensors.

3) Signal Analysis Methodology: We developed a two-pronged approach to analyze the effect of speech on the motion sensors. In the initial step, we analyze the motion sensor signal in the *frequency domain* to look for footprints indicating the presence of speech. If the frequency spectrum shows such an evidence, techniques proposed in [1] and [2] could be used to further classify, recognize or reconstruct the speech signal (such classification is beyond the scope of our paper). If the frequency spectrum is unable to show any evidence, we analyze the signal in *time domain* to look for effects of speech on motion sensors.

Frequency Domain Analysis: To perform the analysis of the motion sensor behavior in presence of speech in the frequency domain, we record speech through the motion sensors and plot the spectrum of the observed signal. We perform similar procedure as prescribed in [1] by playing a 280Hz tone and a multi-tone (consisting of signals having frequencies between 130Hz and 200Hz) from a device for machine-rendered scenarios. Since motion sensors have low sampling rates, the observed frequency range is limited. In case of gyroscope, the sampling rate is 200Hz so observable frequency is limited to 100Hz. Due to this behavior, we depend upon aliasing effect to detect the effects due to the played sound on the spectrum at sub-100Hz frequency range [1].

Time Domain Analysis: In order to measure the presence of a noticeable response of the motion sensors against speech in time domain, we need to compare their behavior in the presence and absence of speech signals. This requires creating two (nearly) identical environments for all the previously described scenarios where one environment contains speech and the other environment is devoid of speech. Placing identical sensors in both environments and measuring their response would accurately determine the susceptibility of motion sensors against human speech. However, creating acoustically identical environments may prove to be a challenge where all parameters like temperature, humidity, pressure, the material and the design of the environment need to be same and constant throughout the experimental phase.

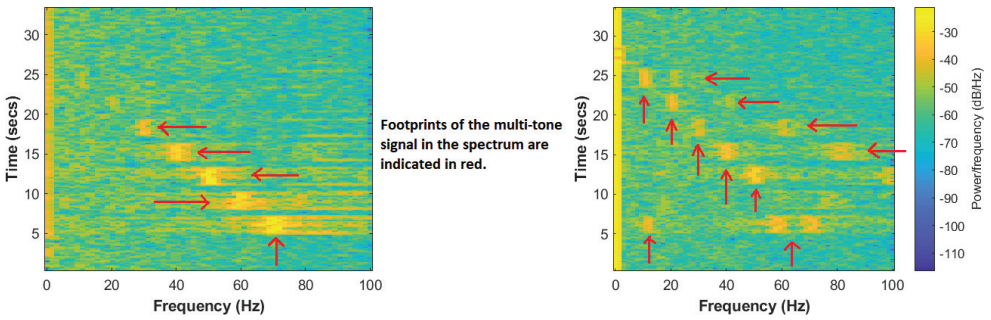
An anechoic chamber as suggested in [1] may be deemed suitable for creating identical acoustic chambers. However, in our work, we circumvent this challenge by performing the normal experiment with human speech immediately followed by a control experiment with no speech, under normal room conditions (in a quiet laboratory room inside university building). If we do not allow any sudden and significant interference (acoustic or vibration) in the environment between the experiments, it should be safe to assume that all the environment variables remained almost constant throughout the experiment. This means the experiments were performed under almost similar conditions and the only noticeable effect should be due to the human speech. In that case, our setup would be emulating the behavior of nearly identical acoustic environments, as described previously.

We recorded and analyzed the sensor readings looking for noticeable effect such as increase in sensor values that may indicate towards presence of speech. We observed multiples audio samples from TIDigits speech corpus[24] and concluded that the pronunciation of a single digit in the corpus took no more than one second. The effect of speech on motion sensor readings lasts for around 0.5 seconds meaning for a sampling frequency of 200Hz (as deployed by the motion sensors), this time duration equates to 100 samples. Thus, we windowed each recorded sample using a window size of 100 samples with an overlap of 50 samples. In each window, we calculated the maximum range achieved by the sensor, which will give us an idea of the disturbance in the readings. If speech signal were strong enough to affect the sensors, the readings would be much higher thereby producing a higher range (due to sensors recording more motion data) when compared to sensor value ranges observed in a relatively silent environment.

Sensor Reading Application: We used the Android application available at [25] to capture sensor readings but modified the source code to include accelerometer sensor readings.

V. Sensor Behavior under Quiet Conditions

Before studying the effect of speech on motion sensors, we observe motion sensors' behavior under ambient conditions, i.e., in a quiet environment. This behavior can then be compared against the behavior of motion sensors under the influence of speech with the assumption that the acoustic environment remains the same.



(a) Gyroscope placed on a wooden surface

(b) Accelerometer placed on a wooden surface

Fig. 1: Spectrum of the motion sensors along x axis that shows the effect of low frequencies contained in the multi-tone signal (130-200Hz). The vibrations due to these frequencies are transmitted along the surface to the motion sensor of the smartphone.

A. Experiment Setup

Equipment: For all our experiments, we use Nexus 5 smartphone that contains the 6-axis motion sensor MPU6515 chip designed by Invensense Inc. It combines a 3-axis gyroscope, 3-axis accelerometer along with a Digital Motion Processor in a single chip. The output precision of the readings is 16 bits for both gyroscope and the accelerometer, and offers a programmable full-scale range of $\pm$2g, $\pm$4g, $\pm$8g and $\pm$16g for accelerometer and up to $\pm$2000dps for the gyroscope. Typical resonant frequency for the gyroscope is listed as 27kHz and the sampling frequency ranges from 4Hz to 8000Hz [26]. The sampling frequency for the accelerometer is described as ranging from 4Hz to 4000Hz. Since Nexus 5 operates on Android platform, the sampling rates for the motion sensors are limited to 200Hz.

We also examined few other smartphone motion sensors available in the market. STMicroelectronics' LSM6DS3 motion sensor in Samsung Galaxy S7 offers similar precision and programmable full-scale range for accelerometer and gyroscope sensors as the Invensense motion sensors. Appendix Table I lists some of the common smartphones and the motion sensors used in these devices. From the table, we see that most of the devices are using either Invensense or STMicroelectronic sensors. The output data rate for all the Invensense sensors is similar for both gyroscope and accelerometer except Nexus 4 (MPU-6050) which is using an older chip. Similarly, the typical mechanical frequency for gyroscope for all the Invensense motion sensor chips is similar except for Nexus 4 again for the reasons specified previously. It also seems that STMicroelectronics do not publish the resonant frequency for their gyroscopes in the data sheet. STMicroelectronics motion sensor differs from Invensense mostly in its output data rate for gyroscope and accelerometer. User programmable range is uniform across the vendors. Thus, we believe that the motion sensors used in our experiments cover a general representation of motion sensors in the market.

Location: We recorded motion sensor readings at four different locations (quiet university lab spaces, henceforth denoted as locations 1, 2, 3, and 4) that acted as near quiet environment. At each location, the ambient noise level was 50 dB. Two of the locations were office rooms inside two different graduate student labs and the rest were conference rooms within the lab spaces. The rooms were devoid of any human presence and the only possible source of noise was the air conditioning vents installed in the ceiling. The recordings were done for an hour to get an estimate of motion sensors' behavior at rest in a quiet environment. The phone was placed on a flat tabletop recording the sensor readings through the sensor reading application.

B. Results

We divided sensor data into samples of length 10 seconds each and calculated the maximum range for each sample. We averaged the obtained maximum range values of sensor readings to get same number of representative samples of sensor readings as the number of samples collected in our subsequent experiments. We plot these representative samples against samples taken under various scenarios (Section IV-B) to analyze sensor behavior under the influence of speech.

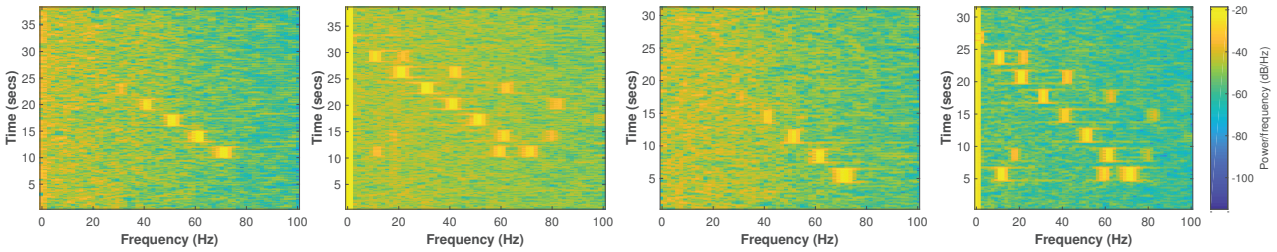
VI. Sensor Behavior against Speech Signals

In this section, we analyze the behavior of motion sensors in the presence of speech. We construct the scenarios described in Section IV and report the results that will help us determine which scenarios are most susceptible to an acoustic side channel attack through motion sensors.

A. Setup Information

Equipment: We use the same device, Nexus 5 from the previous section, where it was used to record the behavior of motion sensors in a quiet environment. For producing machine-rendered speech, we use Logitech Z323 speakers with a frequency response of 55Hz-20kHz that consists of two satellites and a subwoofer (18 watts; 100Hz). Generation of speech signals through smartphone speakers, a recreation of the scenario depicted in [2], was done by iPhone 4S. We also used Thinkpad W530 as the laptop speaker.

Word Data Set: We use the single digit pronunciations from the speech corpus provided at [24] that is a subset of the



(a) Multi-tone on gyroscope 1ft (b) Multi-tone on accelerometer 1ft (c) Multi-tone on gyroscope 4ft (d) Multi-tone on accelerometer 4ft

Fig. 2: Spectrum of the motion sensors readings over different distances in presence of multi-tones (130-200Hz). Bright yellow spots with intensity -20dB/Hz denote the footprints of the multiple frequencies contained in the signal affecting the sensor readings.

TIDIGITS corpus. The speech dataset consists of 5 male and 5 female speakers who perform a single digit pronunciation ("zero" to "nine" with an extra word "oh") which is repeated twice by each speaker. As noted in [1], low sampling frequency restriction on the motion sensors, make it hard to perform speaker-independent speech recognition. Hence, it is reasonable to use a limited dictionary for speaker recognition (such as [24] containing speech of digits) that could still leak confidential information that contains numbers such as social security and credit card numbers, birth dates, PIN etc.

B. Motion Sensors vs. Loudspeaker

In the loudspeaker setup, we test the effect of speech produced by loudspeakers on the motion sensors of a smartphone. The smartphone may reside on same surface as the loudspeaker or on a different surface that is not in physical contact with the loudspeaker.

1) The Loudspeaker-Same-Surface Scenario: We first test the behavior of motion sensors against low frequency tones as a precursor to speech. We recreate the experimental setup from [1] where a 280Hz tone and a multi-tone, consisting of frequencies 130-200Hz, was used. The smartphone is kept on the same surface as the loudspeaker that plays the tones (Appendix Figure 1). We test four different surfaces of which three were wooden desks of varying width and one was a plastic tabletop.

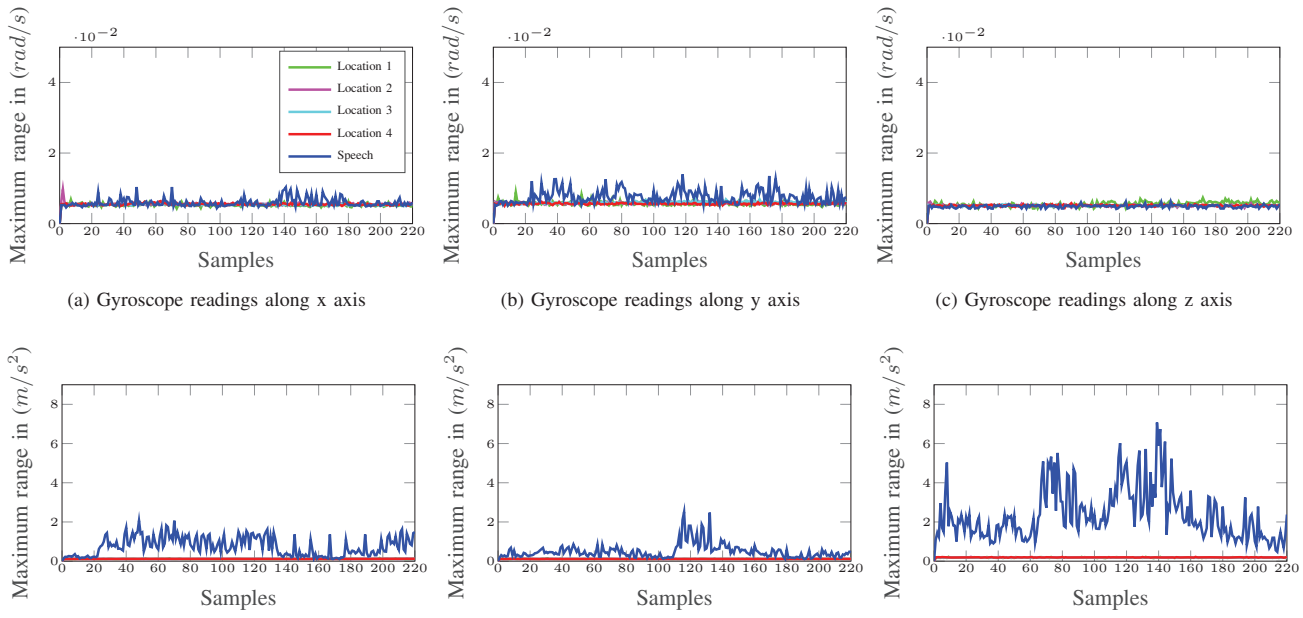
The frequency analysis from our initial experiments showed that playing a 280Hz tone did not affect the gyroscope even when the sound pressure level reached 92db on the wooden surfaces but had some effect on the plastic surface at 92db. At a sound pressure level of 102db, it affected the gyroscope on all the surfaces except one where no effect was observed on gyroscope at 102db. The accelerometer, in contrast, was affected on all the surfaces even at a volume of 72db. When the multi-tone was played, we observed that all the surfaces produced a pronounced effect on both gyroscope and accelerometer when the sound pressure level reached 92db. Figure 1 shows the spectrum for gyroscope and accelerometer for a wooden surface along x axis. The gyroscope was affected along x and y axis of rotation while the accelerometer showed the effect at x, y and z axis of rotation.

We tested Loudspeaker-Same-Surface scenario over varying distance to observe the behavior of motion sensors when the smartphone is placed at different distances from the loudspeaker. The phone is placed at different distances of 1ft, 2ft, 3ft, 4ft, and 5ft with the audio level at the source being kept constant at 92db. The resulting frequency spectrum plots for distance 1ft and 4ft (Figure 2) show the captured signal for the gyroscope (along x axis of rotation) and accelerometer (along x axis). We observe similar intensity of the signal footprint within the small range of our tested distance. This observation indicates that the audio signal may still be captured by the motion sensors even if the motion sensor is not placed close to the loudspeaker. This behavior also indicates that a scenario where loudspeaker and the smartphone reside on same surface such as a conference table, it is possible for motion sensors of the smartphone to get affected due to speech from loudspeakers within the tested distance range.

Effect of Speech: To test the effect of speech on motion sensors, we put the smartphone on the same tabletop (surface) as the loudspeaker and used the word data set [24] as described in Section IV. We set the volume of the loudspeaker to be at the maximum value (99db) to achieve the most response in the motion sensor readings. The surface used was one of the surfaces that had showed effects of 280Hz tone and multi-tone at 90db. The resulting setup is depicted in Appendix Figure 1.

Frequency Domain Analysis: We plot the recorded signal from gyroscope and accelerometer in the frequency domain that are depicted in Appendix Figure 4a (x axis rotation) and Figure 4b (x axis), respectively. Similar frequency spectrums were found for y and z axis rotation for the gyroscope and y and z axis for the accelerometer. From the spectrum, we can see a noticeable footprint for the speech signal on the accelerometer spectrum (around the 3 second mark) that is absent in the gyroscope spectrum. Thus, accelerometer seems to be more sensitive to conductive vibrations from the surface than the gyroscope. Even though, in our set-up we did not observe any noticeable effect on gyroscope, the fact that multi-tone showed an effect on gyroscope indicates that such an effect may also exist for speech signals, as shown in [1].

Time Domain Analysis: The results from time domain analysis



(a) Gyroscope readings along x axis

(b) Gyroscope readings along y axis

(c) Gyroscope readings along z axis

(d) Accelerometer readings along x axis

(e) Accelerometer readings along y axis

(f) Accelerometer readings along z axis

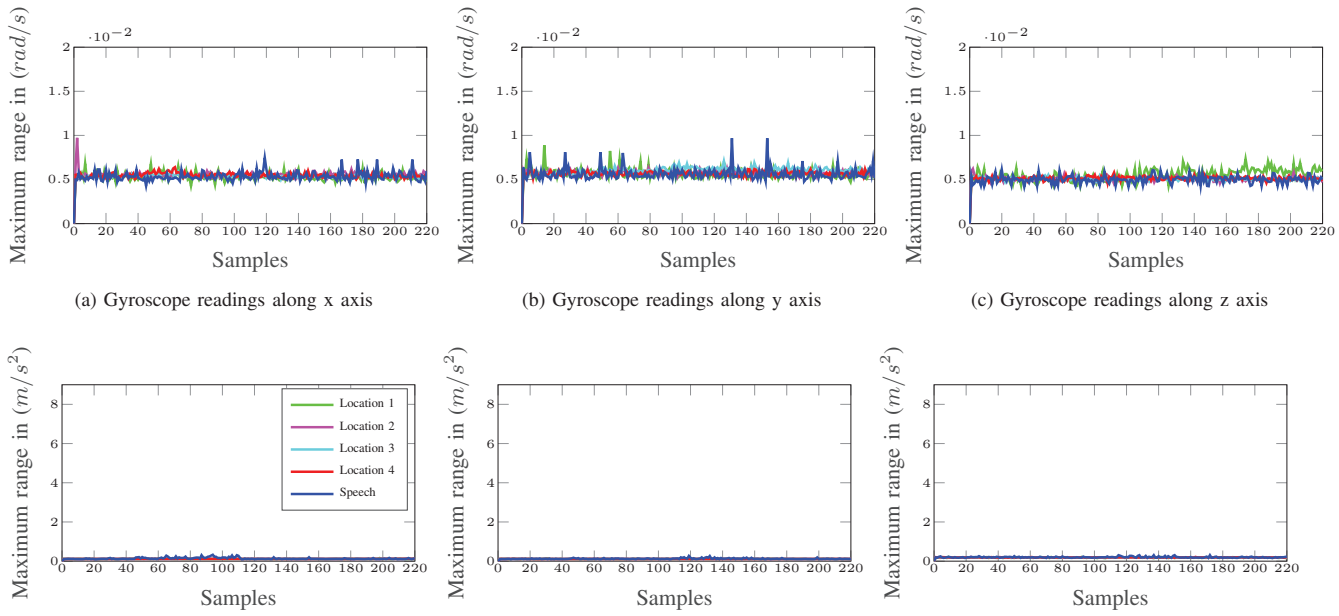
Fig. 3: Comparison of sensor behavior under ambient locations and in presence of speech in the Loudspeaker-Same-Surface scenario. Maximum variance in sensor readings (in absence of speech) at quiet locations 1, 2, 3, 4 is plotted along side maximum variance in sensor readings (in presence of speech) to determine the effect of speech on sensors. Due to surface vibrations from loudspeaker, there is a noticeable effect on accelerometer readings that pushes the blue line plot significantly higher than the line plots of quiet locations (denoted by green, magenta, cyan, and red line plots).

are shown in Figure 3. The gyroscope readings for x axis (Figure 3a) show that maximum variation in sensor readings in presence of speech are comparable to readings taken in absence of speech at quiet locations (between 0.004 and 0.010 rad/s). Maximum variation in readings for y and z axis follow similar pattern (Figure 3b and 3c) falling between 0.005 and 0.012 rad/s, and between 0.004 and 0.007 rad/s. The accelerometer readings in Figure 3d, for the x axis, show that the maximum variation (between 0.104 and 2.066 m/s^2) in presence of speech is higher than in absence of any speech at the four locations (at 0.107 m/s^2). Maximum variations shown on y axis (Figure 3e) follow similar pattern (between 0.113 and 2.511 m/s^2) for speech and 0.111 m/s^2 without speech, while maximum variations in readings along z axis (Figure 3f) are much more higher than the readings along x and y axis. The readings are between 0.555 and 7.073 m/s^2 in presence of speech while they remains the same (around 0.111 m/s^2) in absence of speech.

2) The Loudspeaker-Different-Surface Scenario: For the *Loudspeaker-Different-Surface* scenario, we put the smartphone on a different surface from the loudspeaker and played the word list from the word data set [24] as described in the previous section. We set the volume of the loudspeaker to be at the maximum value (99db) to achieve the highest possible response in the motion sensor readings. The resulting setup is depicted in Appendix Figure 2.

Frequency Domain Analysis: We analyzed the signal from gyroscope and accelerometer readings in the frequency domain. The resulting spectrum is depicted in Figure 4c and Figure 4d, respectively. The plotted gyroscope readings are along x axis of rotation for the gyroscope and along x axis for the accelerometer. We looked for signs of speech around 3.35 second mark that denotes the beginning of speech as per the microphone recording. Both the spectrum figures seem to be devoid of any noticeable footprint of the speech signal around the intended time mark. This leads us to believe that speech signals traveling through air may have no noticeable effect on the motion sensors as per the frequency domain analysis.

Time Domain Analysis: In addition to frequency domain, we also analyzed the readings in time domain as per our measurement metrics and the results are shown in Figure 4 for gyroscope and accelerometer. The gyroscope readings along x and z axis in Figure 4a and 4c show that sensor behavior is similar in presence and absence of speech (maximum range varies between 0.004 and 0.008 rad/s). The readings for y axis in Figure 4b show that the sensor reading with and without speech have the variation for maximum range around 0.005 and 0.009 rad/s. We observe two spikes for speech crossing 0.009 rad/s but there also exists two green spikes around the same value that indicate that this behavior is also displayed sometimes in absence of speech. For accelerometer, the maximum variation in readings with or without speech



(a) Gyroscope readings along x axis

(b) Gyroscope readings along y axis

(c) Gyroscope readings along z axis

(d) Accelerometer readings along x axis

(e) Accelerometer readings along y axis

(f) Accelerometer readings along z axis

Fig. 4: Comparison of sensor behavior under ambient locations and in presence of speech in the Loudspeaker-Different-Surface scenario. Maximum variance in sensor readings (in absence of speech) at quiet locations 1, 2, 3, 4 is plotted along side maximum variance in sensor readings (in presence of speech) to determine the effect of speech on sensors. The blue line plot depicting maximum variance in sensor readings in presence of speech for the given scenario more or less follows similar pattern as the line plots for quiet location indicating a possible lack of any observable effect on motion sensors due to speech.

falls between 0.089 to 0.328 m/s^2 for x axis. For y axis, this value is between 0.108 and 0.268 m/s^2 while the variations for maximum range in absence of speech are contained around 0.218 m/s^2. The sample readings follow the same variation pattern in presence of speech as in absence of speech for z axis falling between 0.178 to 0.305 m/s^2.

C. Motion Sensors vs. Laptop Speakers

For the *Laptop-Same-Surface* scenario, we use laptop speakers instead of loudspeakers. A laptop's speaker is less powerful than a loudspeaker especially in reproducing low frequency sounds accurately. In our experiments, we used Macbook Air 2013 model laptop and IBM Thinkpad W530 laptop to test the effect of speech signals generated by their speakers on the motion sensors of a smartphone placed nearby on the *same surface*. We chose surface 3 as it was able to show response to 280Hz beginning at 92db which the wooden surfaces were unable to show. We set the volume level in the laptops at their maximum value to induce most response from the motion sensors. Since sound pressure level also depends upon the generated signal, we report the sound pressure level for each tested signal accordingly.

Recreating the steps used in the loudspeaker setup (Section VI-B), we first test the response of the motion sensors against a 280Hz tone and a multi-tone signal (130-200Hz) generated by Macbook Air. The spectrum of the frequency domain for gyroscope and accelerometer reveal no response for the 280Hz tone. For the multi-tone signal, the accelerometer shows a faint response along z axis as per Figure 5. Contrasting the result with the results obtained under loudspeaker setup in Section VI-B1, we find that the loudspeaker was able to correctly reproduce multi-tone signal that affected both gyroscope and the accelerometer at 92db. The speaker from laptop was only able to output the multi-tone signal at an average sound pressure level of 67db and hence was unable to produce any noticeable effect on gyroscope and a slight effect on accelerometer along z axis.

Thinkpad W530's speakers performed worse and were unable to output any audible sound for the multi-tone signal with the average sound pressure level of 52db and no response was produced on the frequency spectrum of either the gyroscope readings or the accelerometer readings. Since laptop speakers are not designed to produce enough bass effect, low frequency sounds tend to get lost or distorted when played via laptop speakers. In addition, due to limited sound volume, they are unable to induce vibrations in the surface, that are powerful enough to affect the motion sensors.

Similar behavior is observed when we played male and female speech samples from the word data set as detailed in Section VI-A. Both male and female speech samples could only be generated at a sound pressure level of 80db and 71db respectively and were unable to generate any response

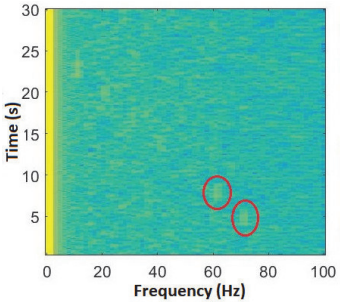


Fig. 5: Spectrum of accelerometer readings (z axis) for multi-tone signal (130-200Hz) generated from laptop. Faint energy signatures (circled in red) can be seen on the spectrum indicating vibration effect produced due to the audio signal.

in the motion sensors. Thus, we believe that a limit to the actual sound volume produced and the inability to correctly reproduce low frequency sounds severely limits the capability of the laptop speakers to affect the motions sensors in a smartphone. Since our time domain analysis in the loudspeaker setup failed to reveal any observable effect on the gyroscope for the speech signals, laptop speakers being less powerful may similarly be unable to reproduce any observable response in the gyroscope sensor. Since accelerometer did show some hints of the multi-tone signal (Figure 5) on frequency spectrum, a more sophisticated frequency domain analysis [1] could be used for further investigation of the presence of speech signal in accelerometer readings.

Based upon our findings in *Laptop-Same-Surface* scenario, we believe that *Laptop-Different-Surface* scenario would produce similar results. In *Laptop-Different-Surface* scenario, the laptop is placed on a different surface from the smartphone and the laptop generates speech signals. Since there exists an air gap between the sensors and the source of speech signals in *Laptop-Different-Surface* scenario in contrast to *Laptop-Same-Surface* scenario, attenuation would be greater for the speech signals. In addition, the inability of Macbook Air speakers to produce the multi-tone signal above 67db, and male (female) speech signals above 80db (71db), it makes it unlikely for laptop produced speech signals to have any noticeable effect in *Laptop-Different-Surface* scenario (due to the fact that no such effect was observed in *Loudspeaker-Different-Surface* scenario even with loudspeakers producing sound at 99db). For laptops that are able to output high quality loud sounds, comparable to a loudspeaker, we believe it would behavior in a similar manner as Loudspeaker-Same-Surface scenario if placed on same surface or Loudspeaker-Different-Surface scenario when placed on different surface.

D. Motion Sensors vs. Phone's Speakers

To examine the behavior of motion sensors against phone speakers, we used an iPhone^{TM}4S phone speaker and played the word list from the word data set [24], as described in the previous section. We set the volume of the phone speaker at the maximum value to achieve most response in the motion sensor readings. The resulting setup is similar to *Loudspeaker-Different-Surface* scenario and is depicted in Appendix Figure 3. It is a recreation of the experimental setup of [2].

Frequency Domain Analysis: The analysis of the motion sensor signals shows no visible indication of speech signal footprints in the frequency spectrum (Appendix Figure 5). This behavior follows the behavior of *Loudspeaker-Different-Surface* scenario from Section VI-B2. Since phone speakers are considerably less powerful than a loudspeaker system, we expect them to perform worse than loudspeakers.

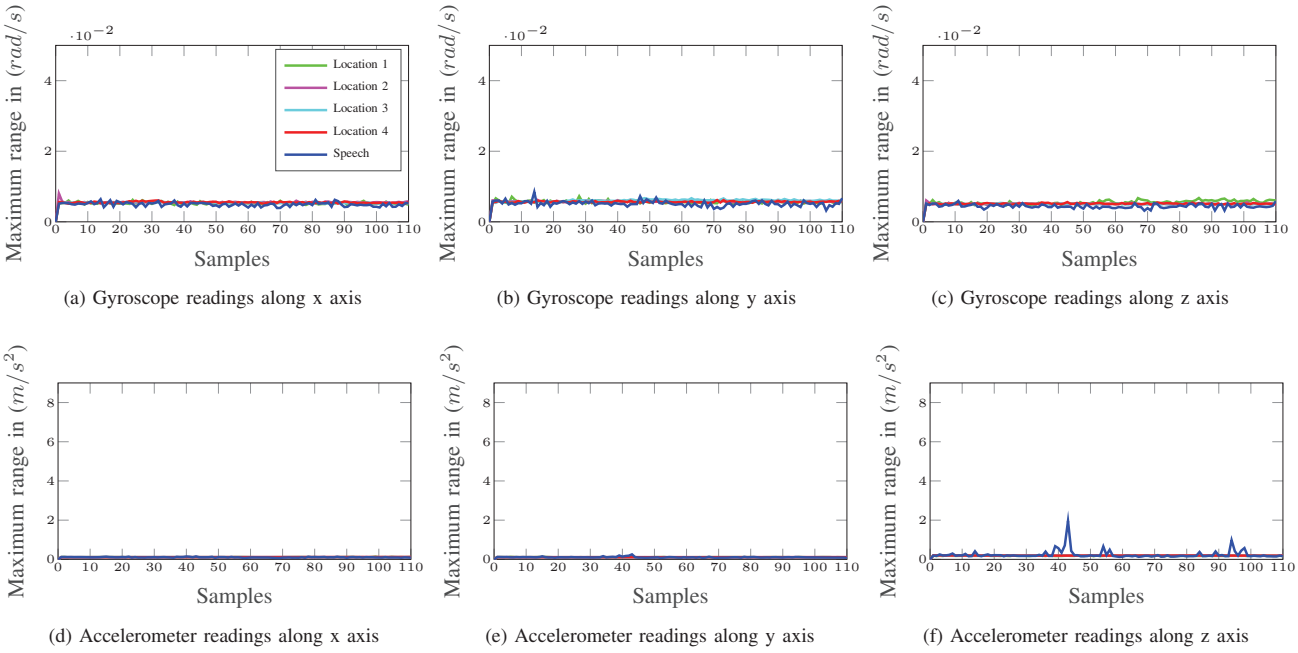
Time Domain Analysis: The results of time domain analysis of the sensor readings in presence of phone speakers on a different surface are shown in Appendix Figure 6. We observed that the maximum variation in gyroscope readings hovered around 0.006 rad/s for x axis (Appendix Figure 6a) and around 0.005 rad/s for y and z axis (Appendix Figure 6b and 6c). The accelerometer analysis shows us that maximum variations in its readings are consistent in presence and absence of speech at around 0.1 m/s^2 for x, y and z axis (Appendix Figure 6d, 6e and 6f). These comparisons lead us to believe that phone speakers may not have any profound effect on smartphone sensors.

E. Motion Sensors vs. Human Speech

1) The Human-Normal Scenario: We recruited 10 human subjects (ages 20-40; 5 males, 5 females; graduate and above education level) for our study that was approved by our University's IRB. The participation was voluntary and the participants could withdraw any time. Our sample size, while small in size, is adequate for judging motion sensor behavior in presence of live human speech as we also provide for two scenarios (normal and loud) for covering possible cases that may occur in real world. All the subjects were healthy (reported on their own accord) and did not suffer from any ailments that affected their vocal chords. Each subject was asked to recite the list of words that is used in [24]. The subjects were seated in a quiet room on a chair, and were not in physical contact with the table. The smartphone was placed face-up on the table and was running the same application as in previous experiments. The subjects were asked to take extra care not to touch or move the table during the entire duration of the experiment as to not affect the motion sensors. The subjects were asked to go through the word list by speaking each word in a normal conversational tone into the phone at a close distance of 10cm.

Frequency Domain Analysis: In the frequency domain, the spectrum of the recorded human speech by gyroscope does not reveal any effects on the sensor readings along x, y and z axis of rotation as per Appendix Figure 8a, 8b and 8c. Similar pattern was found in the accelerometer readings as per Figure 8d, 8e and 8f.

Time Domain Analysis: The observations for *Human-Normal* scenario are depicted in Figure 7 for gyroscope and accelerometer. We observe that maximum observed range in gyrosocpe readings during speech is approximately 0.004 and 0.006 rad/s for x axis (except outlier samples 17 that is around



(a) Gyroscope readings along x axis

(b) Gyroscope readings along y axis

(c) Gyroscope readings along z axis

(d) Accelerometer readings along x axis

(e) Accelerometer readings along y axis

(f) Accelerometer readings along z axis

Fig. 6: Comparison of sensor behavior under ambient locations and in presence of live human speech (loud). Maximum variance in sensor readings (in absence of speech) at quiet locations 1, 2, 3, 4 is plotted along side maximum variance in sensor readings (in presence of speech) to determine the effect of speech on sensors. The blue line (maximum variance in presence of speech) closely follows rest of the lines (that represent maximum variance in absence of speech) indicating similar behavior of sensors in quiet locations and against loud human voice.

0.009 rad/s), 0.003 and 0.009 rad/s for y axis, and 0.003 and 0.007 rad/s for the z axis (Figure 7a, 7b and 7c). This behavior follows similar to sensor readings taken in absence of speech. For accelerometer, these values are approximately 0.1 m/s^2 for x and y axis, and 0.18 m/s^2 for z axis (Figure 7d, 7e and 7f).

2) The Human-Loud Scenario: This scenario is similar to *Human-Normal* scenario with same subjects and same equipment and location being used. The only change in this scenario from *Human-Normal* scenario is that the subjects were asked to shout as loud as they could instead of speaking normally into a smartphone placed on the table. The word list is kept same and the instructions to avoid any physical contact with the table were followed.

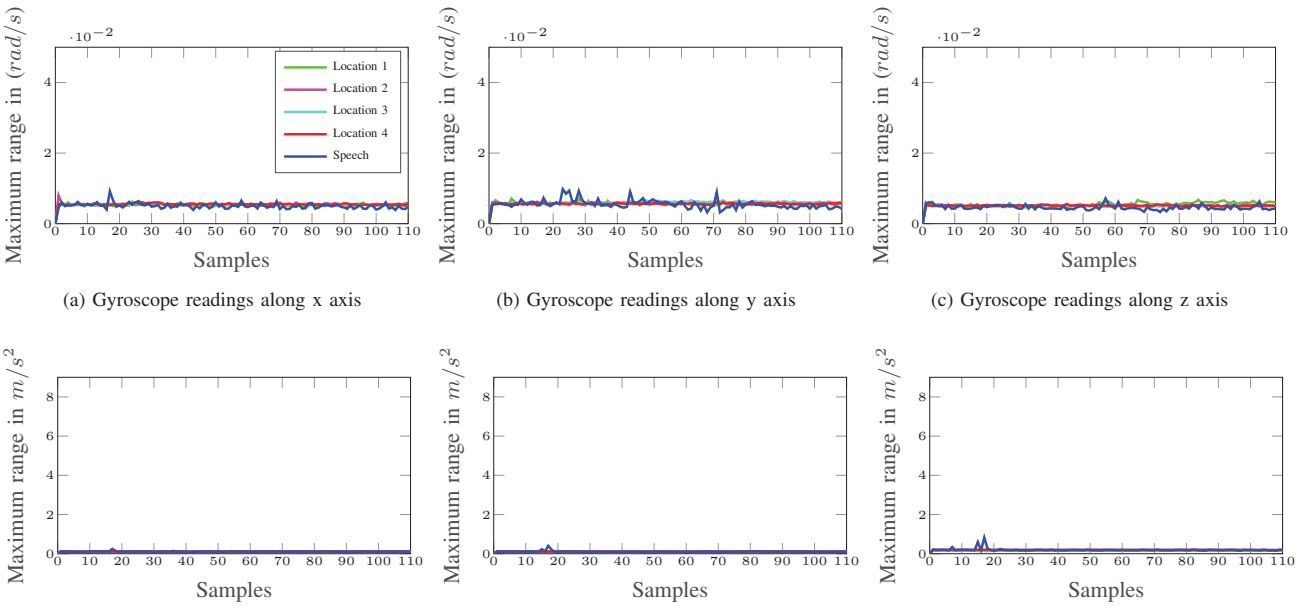
Frequency Domain Analysis: The spectrum of recorded human speech, spoken as loud as possible, by gyroscope does not reveal any effects on the sensor readings along x, y and z axis of rotation as per Appendix Figure 7a, 7b and 7c. We find similar behavior in the accelerometer readings as per Figure 7d, 7e and 7f.

Time Domain Analysis: We calculated maximum observed range for the sensor readings in *Human-Loud* scenario. Our findings, as shown in Figure 6a, 6b and 6c, indicate that the metrics for gyroscope are around 0.005 rad/s for x axis. The values for y and z axis are around 0.005 rad/s. Overall, these numbers are very similar to our observations under speech free environment at locations 1, 2, 3 and 4 in the plot. The values from accelerometer are around 0.1 for x and y axis (Figure 6d and 6e) and 0.2 for z axis as shown in Figure 6f (with samples 43, 94 and 97 being the most obvious outliers possibly due to physical interaction of the user with the phone). These results for accelerometer indicate similar behavior as in speech free environment depicted by readings from Location 1, 2, 3 and 4 in Figure 6 and hence serve to consolidate our conclusion from the frequency domain analysis.

VII. Summary and Further Insights

Recalling the three factors (source of speech, audio transfer medium and sound pressure level) that are important in deciding whether a noticeable effect would be produced by the speech on the sensors, we pick each of the tested scenarios, summarize the obtained results and present the insights learned from our analysis (summary of results is depicted in Table I).

Motion Sensors against Loudspeaker: The obtained results from this setup indicate that motion sensors could possibly be affected by the speech signals rendered by loudspeaker, when the loudspeaker setup and the motion sensors are sharing the same surface. Since the smartphone is in contact with same surface as the loudspeaker setup, the conductive vibrations through loudspeaker setup travel through the shared surface to reach motion sensors. The amount of dampening of conductive



(a) Gyroscope readings along x axis

(b) Gyroscope readings along y axis

(c) Gyroscope readings along z axis

(d) Accelerometer readings along x axis

(e) Accelerometer readings along y axis

(f) Accelerometer readings along z axis

Fig. 7: Comparison of sensor behavior under ambient locations and in presence of live human speech (normal). Maximum variance in sensor readings (in absence of speech) at quiet locations 1, 2, 3, 4 is plotted along side maximum variance in sensor readings (in presence of speech) to determine the effect of speech on sensors. The blue line (maximum variance in presence of speech) closely follows rest of the lines (that represent maximum variance in absence of speech) indicating similar behavior of sensors in quiet locations and against normal human voice.

TABLE I: Summary of the impact of speech on motion sensors. Only scenarios studied in prior work were: *Loudspeaker-Same-Surface* against gyroscope [1] and *Phone-Different-Surface* against acceleroemter [2].

Scenario	Gyroscope affected	Accelerometer affected
Loudspeaker-Same-Surface	*possibly**	✓
Loudspeaker-Different-Surface	×	×
Laptop-Same-Surface	×	*possibly**
Phone-Different-Surface	×	×
Human-Normal	×	×
Human-Loud	×	×

*Sophisticated analysis in the frequency domain using machine learning may be able to reveal the impact of speech, as shown by [1] in the Loudspeaker-Same-Surface scenario.

vibrations depends on the surface material though no noticeable effect was seen in our experiments up to a distance of 4 feet on a wooden table surface. Similar conductive vibration effects have also been reported in another attack vector [3] where vibrations associated with user's key presses, travel over to a smartphone close by kept on the same surface as the computer.

In the scenario where the loudspeaker and the smartphone did not share a surface, in contrast, the obtained gyroscope readings show similar behavior as when taken under quiet conditions, that lead to the conclusion that gyroscope sensor remains unaffected by the loudspeaker in this scenario. Similar results from the examination of the accelerometer readings point that the motion sensors possibly remain unaffected in this scenario. Since this scenario was designed to have no direct or indirect contact between the loudspeaker and the smartphone (except for perhaps the ground), it removes the medium of transfer for the conductive vibrations while the acoustic vibrations can still travel through the air. Thus, we believe that it were the conductive vibrations that affected the motion sensor readings. In addition, we believe that acoustic vibrations are unable to produce a significant impact on the motion sensors as observed by the lack of a response in the *Loudspeaker-Different-Surface* scenario. Therefore, we can also assume that the behavior observed in [1] may have been due to (indirect) conductive vibrations as opposed to (direct) acoustic vibrations.

Motion Sensors against Laptop Speaker: In this scenario (*Laptop-Same-Surface*), we observed that only accelerometer was slightly affected by low frequency tones but speech signals rendered by laptop speakers were not powerful enough to induce a response in both gyroscope and accelerometer.

Motion Sensors against Phone Speaker: In this scenario (*Phone-Different-Surface*), we observed no significant impact on gyroscope and accelerometer due to the speech signals transmitted from the phone speaker. Since this scenario is a weaker setting than the loudspeaker setup due to the fact that

phone speakers lack the loudness and the richness of the sound produced by a loudspeaker, this result seems to indicate that the motion sensors may remain unaffected by speech signals produced from smartphone speakers.

Motion Sensors against Normal Human Voice: The results from this scenario (*Human-Normal*) indicate that there is minimal variation in the sensor readings when compared against the sensor readings taken in absence of speech, leading us to believe that human speech in normal conversational tone may be unable to produce significant response in the motion sensor readings.

Motion Sensors against Loud Human Voice: This scenario (*Human-Loud*) involved much louder human speech and showed similar behavior for the motion sensors that was observed in the previous scenario (*Human-Normal*). The results leads to the notion that even loud human speech may not be strong enough on its own to have a significant impact on the motion sensors that are embedded in the smartphones.

All these results seem to indicate that direct acoustic vibrations are unable to affect the motion sensors while traveling through air. However, it is possible for other sounds such as high frequency audio signals to affect gyroscope and accelerometer as shown in [12], [13], [14], [15]. The difference lies in the frequencies and power levels of the audio signals used to influence the motion sensors. While the fundamental frequency for speech is in the range 85-180 Hz (male) and 165-255 Hz (female), the audio signal used in prior work has been near the resonant frequencies of the motion sensors and around high sound pressure level of 90dB and above.

Accelerometer vs. Gyroscope: An interesting insight from our experimental results (Table I) is that the accelerometer seems to be more sensitive towards conductive vibrations than the gyroscope. This behavior may be due to the reason that in all our setups, the smartphone was placed on a flat surface. Thus, linear motion along x, y axes may be natural (as captured by the accelerometer), while rotation along x and y axes may be restricted (as captured by the gyroscope) due to the surface on which the smartphone is resting.

VIII. Potential Future Work

We showed that motion sensors seem to get impacted by the speech signal only in certain scenarios depending upon the source of speech generation, the medium of transfer and sound pressure level. For such scenarios, future work may be conducted, applying machine learning methods similar to the work done in [1], in order to detect and classify the speech impact to achieve speaker and gender identification in frequency domain. To further strengthen the role of *conductive vibrations* in the success of attacks that exploit motion sensors for compromising speech privacy, in the threat scenarios examined in this work, such vibrations could be measured via Laser Doppler Vibrometer (LDV). Correlating the surface vibrations with the motion sensor readings would reaffirm the role played by *conductive vibrations* in such attacks.

Increasing the sampling rate of the motion sensors may help them capture and record more information. This effect can be achieved by using motion sensors that have a higher sampling rate, though operating systems could be tempted to limit the sampling rate as a security measure. Another way to increase the sampling rate would be to consider a more relaxed threat model where the adversary has the capability to override the limit imposed by the operating system on the sampling rate of the motion sensors via a covert malware application. It is also possible for future generations of smartphones to have an increased sampling rate for motion sensors for better accuracy. This feature will possibly lead to better accuracy for the adversary in the discussed threat scenarios and hence any such design decisions need to be taken keeping such factors in mind. The use of multiple sensors in a fashion similar to an array of time interleaved data converters (interleaved ADCs) could also be used to artificially ramp up the sampling rate as also suggested in [1], [27]. This line of work would be another interesting item for possible future work.

Further work may also explore other side-channel attacks against motion sensors present in the research literature (e.g. [3], [18], [17], [28], [29], [30]), and how it would affect the attack feasibility if these attacks were to be extended to different novel scenarios, in line with the theme of our work. We believe such a discussion is necessary in order to enrich the threat assessment of side-channel attacks on motion sensors under different setups.

IX. Concluding Remarks

In this work, we conducted a threat analysis of the motion sensors embedded in current smartphone platforms against speech signals. In particular, we examined the possibility of compromising the speech privacy of a user by exploiting the motion sensor data in a covert fashion. We conducted our study covering many possible attack scenarios and analyzed the behavior of the motion sensors under all these scenarios. Taking into account the performed investigation, we reached the conclusion that the threat levels perceived due to motion sensors' recording of speech signals depend on a number of factors and seem mostly due to result of *conductive vibrations* produced by the speech generating device. The impact of the studied threat under the designed scenarios therefore limits itself to only specific settings. Further work must be conducted similar to the threat scenarios designed in this work to assess other computing platforms and paradigms that incorporate motion sensors against speech privacy vulnerability, or lack thereof (e.g., the embedded devices in the IoT space or an aggregation of multiple devices with multiple motion sensors).

Acknowledgment

The authors would like to thank our shepherd Dr. Kevin Fu and anonymous reviewers for their feedback on the paper. We would also like to thank Prakash Shrestha for valuable suggestions on a previous draft of the paper. This work was supported in part by National Science Foundation under the grant NSF CNS-1526524.

REFERENCES

[1] Y. Michalevsky, D. Boneh, and G. Nakibl, "Gyrophone: Recognizing speech from gyroscope signals," in *23rd USENIX Security Symposium (USENIX Security'14)*, 2014, pp. 1053–1067.

[2] L. Zhang, P. H. Pathak, M. Wu, Y. Zhao, and P. Mohapatral, "Accelword: Energy efficient hotword detection through accelerometer," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, 2015, pp. 301–315.

[3] P. Marquardt, A. Verma, H. Carter, and P. Traynor, "(sp)iphone: Decoding vibrations from nearby keyboards using mobile phone accelerometers," in *ACM Conference on Computer and Communications Security (CCS'11)*, 2011.

[4] "Using the gyroscope to record sound without microphone permissions on android," https://www.reddit.com/r/netsec/comments/2e3m5c/using_the_gyroscope_to_record_sound_without/?st=iqa09x7l&sh=e5b7bbd8, 2016.

[5] "The gyroscopes in your phone could let apps eavesdrop on conversations," https://www.wired.com/2014/08/gyroscope-listening-hack/, 2016.

[6] "Can a smartphone gyroscope be an eavesdropping tool?" http://searchsecurity.techtarget.com/answer/Can-a-smartphone-gyroscope-be-an-eavesdropping-tool, 2016.

[7] "Hackers transform a smartphone gyroscope into an always-on microphone," https://www.engadget.com/2014/08/15/gyrophone-eavesdropping-hack/, 2016.

[8] "Your smartphone's gyroscope can be turned into an eavesdropping hacker's microphone," http://www.digitaltrends.com/mobile/your-smartphones-gyroscope-can-be-turned-into-an-eavesdropping-hackers-microphone-privacy, 2016.

[9] "Eavesdropping using smart phone gyroscopes," https://www.schneier.com/blog/archives/2014/08/eavesdropping_u.html, 2016.

[10] "Hacker news," https://news.ycombinator.com/item?id=8178777, 2016.

[11] "How spies could use your phone to eavesdrop on you," http://www.techtimes.com/articles/13147/20140815/how-spies-could-use-your-phone-to-eavesdrop-on-you.htm, 2016.

[12] S. Castro, R. Dean, G. Roth, G. T. Flowers, and B. Grantham, "Influence of acoustic noise on the dynamic performance of mems gyroscopes," in *ASME 2007 International Mechanical Engineering Congress and Exposition*, 2007, pp. 1825–1831.

[13] R. N. Dean, G. T. Flowers, A. S. Hodel, G. Roth, S. T. Castro, R. Zhou, A. Moreira, A. Ahmed, R. Rifki, B. E. Grantham, D. A. Bittle, and J. P. Brunsch, "On the degradation of mems gyroscope performance in the presence of high power acoustic noise." in *Industrial Electronics, 2007. ISIE 2007. IEEE International Symposium on*, 2007, pp. 1435–1440.

[14] R. N. Dean, S. T. Castro, G. T. Flowers, G. Roth, A. Ahmed, A. S. Hodel, B. E. Grantham, D. A. Bittle, and J. P. B. Jr, "A characterization of the performance of a mems gyroscope in acoustically harsh environments." *IEEE Transactions on Industrial Electronics*, vol. 58, no. 7, pp. 2591–2596, 2011.

[15] Y. Son, H. Shin, D. Kim, Y. Park, J. Noh, K. Choi, J. Choi, and Y. Kim, "Rocking drones with intentional sound noise on gyroscopic sensors," in *24th USENIX Security Symposium (USENIX Security'15)*, 2015, pp. 881–896.

[16] L. Cai and H. Chen, "Touchlogger: Inferring keystrokes on touch screen from smartphone motion." *HotSec*, vol. 11, pp. 9–9, 2011.

[17] E. Owusu, J. Han, S. Das, A. Perrig, and J. Zhang, "Accessory: password inference using accelerometers on smartphones," in *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*. ACM, 2012, p. 9.

[18] Z. Xu, K. Bai, and S. Zhu, "Taplogger: Inferring user inputs on smartphone touchscreens using on-board motion sensors," in *Proceedings of the fifth ACM conference on Security and Privacy in Wireless and Mobile Networks*. ACM, 2012, pp. 113–124.

[19] L. Cai and H. Chen, *Trust and Trustworthy Computing: 5th International Conference, TRUST 2012, Vienna, Austria, June 13-15, 2012. Proceedings*. Springer Berlin Heidelberg, 2012, ch. On the Practicality of Motion Based Keystroke Inference Attack, pp. 273–290.

[20] E. Miluzzo, A. Varshavsky, S. Balakrishnan, and R. R. Choudhury, "Tapprints: your finger taps have fingerprints," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACM, 2012, pp. 323–336.

[21] "Coriolis force," https://en.wikipedia.org/wiki/Coriolis_force, 2017.

[22] "Voice Acoustics: an introduction," http://newt.phys.unsw.edu.au/jw/voice.html, 2016.

[23] "MPU-9250 Product Specification Revision 1.0," https://store.invensense.com/datasheets/invensense/MPU9250REV1.0.pdf, 2016.

[24] "Clean Digits," http://www.ee.columbia.edu/~dpwe/sounds/tidigits/, 2016.

[25] "Gyromic," https://crypto.stanford.edu/gyrophone/application/gyromic.apk, 2016.

[26] "MPU-6500 Product Specification Revision 1.0," https://store.invensense.com/datasheets/invensense/MPU_6500_Rev1.0.pdf, 2016.

[27] J. Han, A. J. Chung, and P. Tague, "Pitchln: Eavesdropping via intelligible speech reconstruction using non-acoustic sensor fusion," in *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks*, ser. IPSN '17. New York, NY, USA: ACM, 2017, pp. 181–192.

[28] A. J. Aviv, B. Sapp, M. Blaze, and J. M. Smith, "Practicality of accelerometer side channels on smartphones," in *Proceedings of the 28th Annual Computer Security Applications Conference*. ACM, 2012, pp. 41–50.

[29] S. Dey, N. Roy, W. Xu, R. R. Choudhury, and S. Nelakuditi, "Accelprint: Imperfections of accelerometers make smartphones trackable." in *Network and Distributed System Security Symposium (NDSS)*. Citeseer, 2014.

[30] A. Das, N. Borisov, and M. Caesar, "Exploring ways to mitigate sensor-based smartphone fingerprinting," *arXiv preprint arXiv:1503.01874*, 2015.

APPENDIX

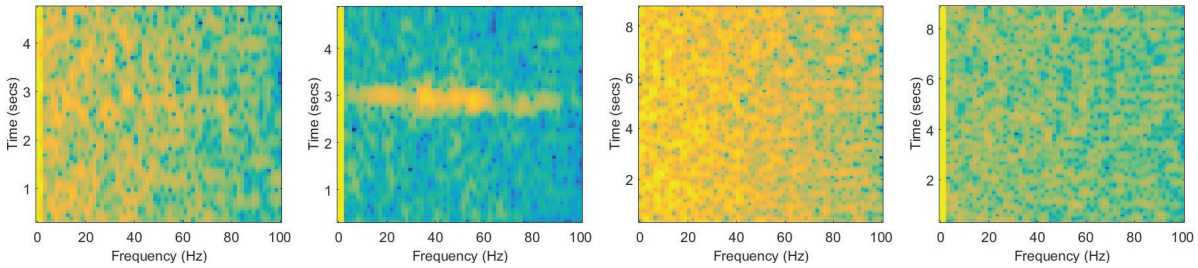
Fig. 1: Experiment setup depicting loudspeaker and the smartphone with embedded motion sensors placed on same surface.

Fig. 2: Experiment setup depicting loudspeaker and the smartphone with embedded motion sensors placed on different surfaces.

Fig. 3: Experiment setup depicting a phone speaker against the smartphone with embedded motion sensors that are placed on different surfaces.

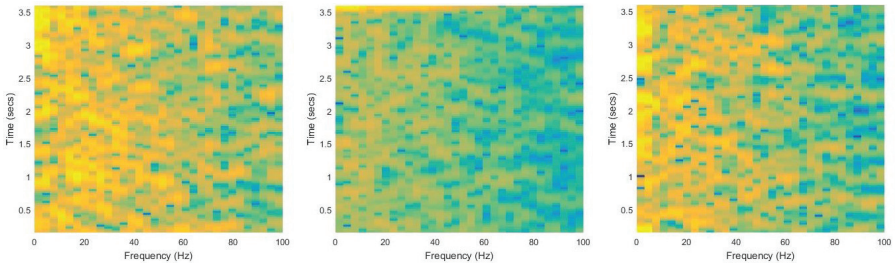
TABLE I: Motion sensors specifications for some popular brands of smartphones (Courtsey: iFixit and Chipworks Inc.) The specifications indicate nearly similar hardware and software features for these motion sensor chips.

Smartphone brand	Vendor	Sensor	Gyroscope			Accelerometer	
			User programmable range	Output data rate	Mechanical frequency	User programmable range	Output data rate
iPhone 6/6s	Invensense	MPU-6700	$\pm 250, \pm 500, \pm 1000, \pm 2000$ dps	4-8000 Hz	27kHz	$\pm 2g, \pm 4g, \pm 8g, \pm 16g$	4-4000Hz
	Bosch	BMA280	N/A	N/A	N/A	$\pm 2g, \pm 4g, \pm 8g, \pm 16g$	2000Hz
iPhone 7	Invensense	ICM-20608-G	$\pm 250, \pm 500, \pm 1000, \pm 2000 dps$	4-8000 Hz	27kHz	$\pm 2g, \pm 4g, \pm 8g, \pm 16g$	4-4000Hz
Samsung Galaxy S7 Edge	STMicroelectronics	LSM6DS3	$\pm 125, \pm 245, \pm 500, \pm 1000, \pm 2000 dps$	12.5-1660 Hz	Not available	$\pm 2g, \pm 4g, \pm 8g, \pm 16g$	12.5-6000 Hz
Samsung Galaxy S6	Invensense	MPU-6500	$\pm 250, \pm 500, \pm 1000, \pm 2000 dps$	4-8000 Hz	27kHz	$\pm 2g, \pm 4g, \pm 8g, \pm 16g$	4-4000Hz
Samsung Galaxy S5	Invensense	MP65M (6500)	$\pm 250, \pm 500, \pm 1000, \pm 2000 dps$	4-8000 Hz	27kHz	$\pm 2g, \pm 4g, \pm 8g, \pm 16g$	4-4000Hz
Nexus 5	Invensense	MPU-6515	$\pm 250, \pm 500, \pm 1000, \pm 2000 dps$	4-8000 Hz	27kHz	$\pm 2g, \pm 4g, \pm 8g, \pm 16g$	4-4000Hz
Nexus 4	Invensense	MPU-6050	$\pm 250, \pm 500, \pm 1000, \pm 2000 dps$	4-8000 Hz	33kHz along x axis; 30kHz along y axis; 27kHz along z axis	$\pm 2g, \pm 4g, \pm 8g, \pm 16g$	4-1000Hz
Samsung Galaxy S3	STMicroelectronics	LSM330DLC	$\pm 250, \pm 500, \pm 2000 dps$	95-760Hz	Not available	$\pm 2g, \pm 4g, \pm 8g, \pm 16g$	1-5376Hz

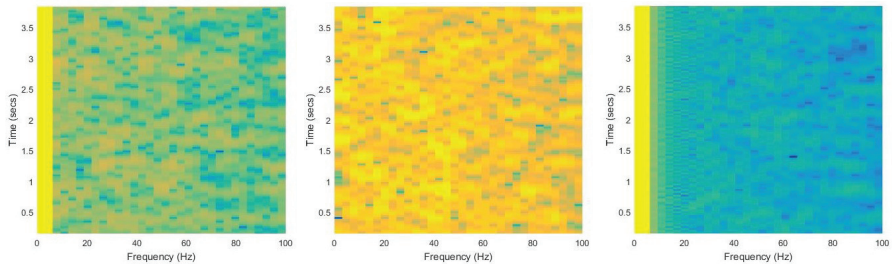


(a) Gyroscope same surface (b) Accelerometer same surface (c) Gyroscope diff. surface (d) Accelerometer diff. surface

Fig. 4: Spectrum showing motion sensor readings along x axis in presence of male voice pronouncing "OH" in loudspeaker setup. There is a lack of noticeable effect on the gyroscope spectrum for both scenarios while the accelerometer spectrum shows the existence of speech for the same surface scenario.

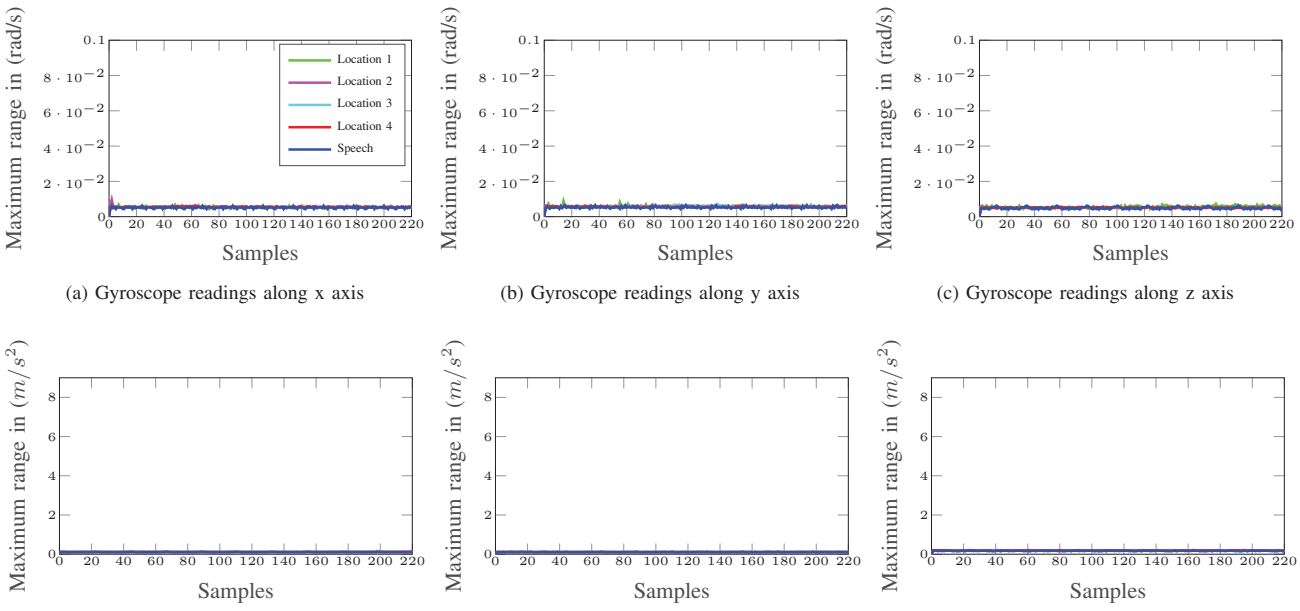


(a) Gyroscope readings along x axis of rotation (b) Gyroscope readings along y axis of rotation (c) Gyroscope readings along z axis of rotation



(d) Accelerometer readings along x axis (e) Accelerometer readings along y axis (f) Accelerometer readings along z axis

Fig. 5: Spectrum of the motion sensors in presence of a phone speaker pronouncing "OH" indicating the lack of any observable effect on motion sensor readings for the setup with phone speaker and a different smartphone containing motion sensor reside on the same surface.



(a) Gyroscope readings along x axis

(b) Gyroscope readings along y axis

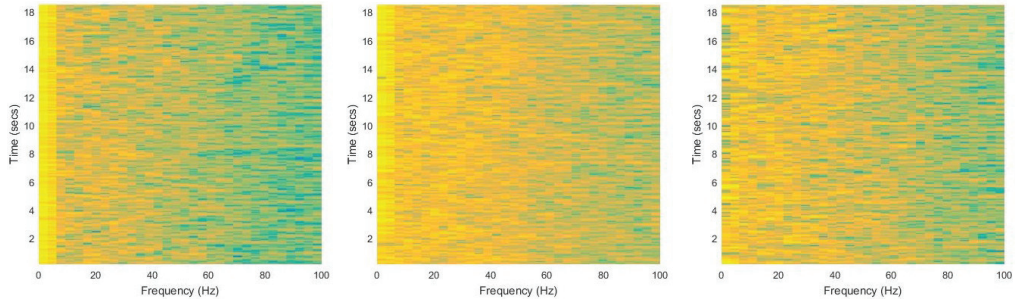
(c) Gyroscope readings along z axis

(d) Accelerometer readings along x axis

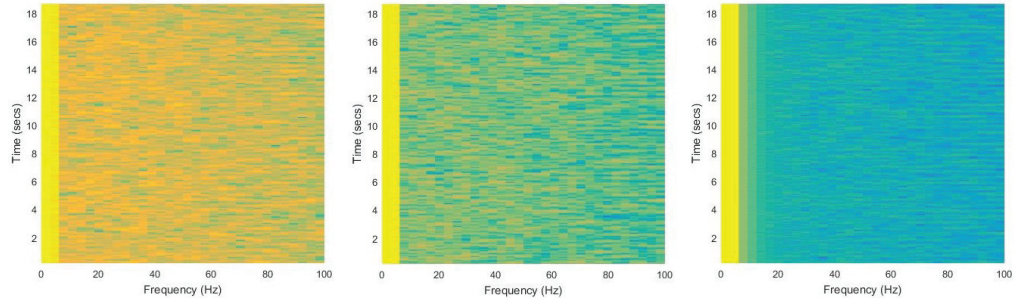
(e) Accelerometer readings along y axis

(f) Accelerometer readings along z axis

Fig. 6: Comparison of sensor behavior under ambient locations and in presence of speech in the Phone-Different-Surface scenario. Maximum variance in sensor readings (in absence of speech) at quiet locations 1, 2, 3, 4 is plotted along side maximum variance in sensor readings (in presence of speech) to determine the effect of speech on sensors. The blue line (maximum variance in presence of speech) closely follows rest of the lines (that represent maximum variance in absence of speech) indicating similar behavior of sensors in quiet locations and under the effect of speech from a phone speaker placed on a different surface.

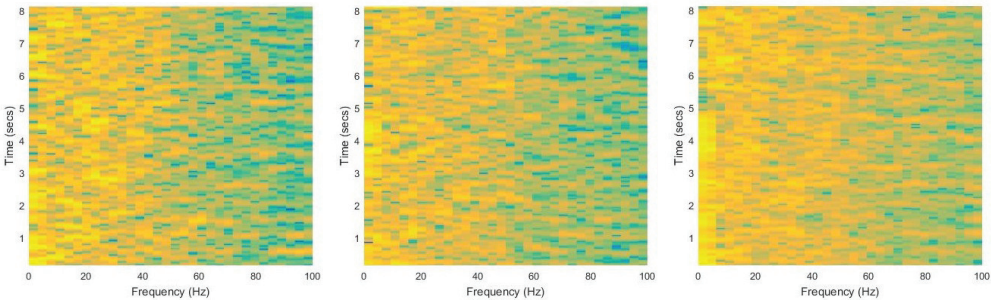


(a) Gyroscope readings along x axis of rotation (b) Gyroscope readings along y axis of rotation (c) Gyroscope readings along z axis of rotation

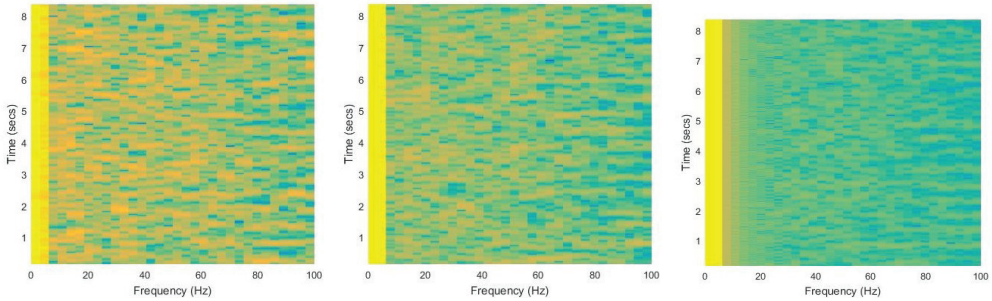


(d) Accelerometer readings along x axis (e) Accelerometer readings along y axis (f) Accelerometer readings along z axis

Fig. 7: Spectrum of the motion sensors in presence of a human speaker (loud) pronouncing "OH" indicating the lack of any observable effect on motion sensor readings for the setup with a smartphone (with motion sensors) under the effect of a human speaker in a loud voice.



(a) Gyroscope readings along x axis of rotation (b) Gyroscope readings along y axis of rotation (c) Gyroscope readings along z axis of rotation



(d) Accelerometer readings along x axis (e) Accelerometer readings along y axis (f) Accelerometer readings along z axis

Fig. 8: Spectrum of the motion sensors in presence of a human speaker (normal) pronouncing "OH" indicating the lack of any observable effect on motion sensor readings for the setup with a smartphone (with motion sensors) under the effect of a human speaker in a loud voice.