NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

AUTHENTICATING A KNOWN USER THROUGH BEHAVIORAL BIOMETRICS USING A SMARTPHONE ACCELEROMETER

by

Patrick W. Jones

December 2017

Thesis Advisor: John McEachen
Co-Advisor: Preetha Thulasiraman

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704–0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** December 2017	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE AUTHENTICATING A KNOWN USER THROUGH BEHAVIORAL BIOMETRICS USING A SMARTPHONE ACCELEROMETER		**5. FUNDING NUMBERS**	
6. AUTHOR(S) Patrick W. Jones			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000		**8. PERFORMING ORGANIZATION REPORT NUMBER**	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A		**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB number 2017.0103-DD-N			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.		**12b. DISTRIBUTION CODE**	

13. ABSTRACT (maximum 200 words)

This thesis investigates the feasibility of authenticating a user through a behavioral biometric signature from smartphone accelerometer data. Using a Samsung Galaxy S7, acceleration in relation to the necessary equilibrium, postural state for a subject to orient a smartphone in order to read a headline article was measured and recorded by the MATLAB Mobile application. Twenty subjects—1 known and 19 unknown—were used in the creation of a MATLAB machine-learning classifier. The classifier accurately distinguished an unknown subject from the known subject. Recommendations for future work include repeating the experiment with the latest smartphone devices as available, incorporating different sensors available to the "MATLAB Mobile App," and introducing noise to spoof the known user.

14. SUBJECT TERMS smartphone, authentication, behavioral biometrics, accelerometer, Android			**15. NUMBER OF PAGES** 51
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01-280-5500 Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

AUTHENTICATING A KNOWN USER THROUGH BEHAVIORAL BIOMETRICS USING A SMARTPHONE ACCELEROMETER

Patrick W. Jones
Lieutenant, United States Navy
B.S., United States Naval Academy, 2010

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
December 2017

Approved by: John McEachen
Thesis Advisor

Preetha Thulasiraman
Co-Advisor

R. Clark Robertson
Chair, Department of Electrical Engineering

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This thesis investigates the feasibility of authenticating a user through a behavioral biometric signature from smartphone accelerometer data. Using a Samsung Galaxy S7, acceleration in relation to the necessary equilibrium, postural state for a subject to orient a smartphone in order to read a headline article was measured and recorded by the MATLAB Mobile application. Twenty subjects—1 known and 19 unknown—were used in the creation of a MATLAB machine-learning classifier. The classifier accurately distinguished an unknown subject from the known subject. Recommendations for future work include repeating the experiment with the latest smartphone devices as available, incorporating different sensors available to the "MATLAB Mobile App," and introducing noise to spoof the known user.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

DARPA	Defense Advanced Research Projects Agency
DOD	Department of Defense
FIR	Finite ImpulseResponse
PIN	Personally Identifiable Number
SGO	Savitzky-Golay
SOCOM	Special Operations Command

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

Thank you Professors McEachen and Thulasiraman for your continued support and guidance during the process of completing this thesis. Thank you to my tribe within the electrical engineering cohort that made the journey not only possible, but more importantly, one of which I will cherish with fond memories.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND AND MOTIVATION

Smartphones are becoming more common and available than ever. Currently, 77% of all adults within the United States report owning a smartphone [1]. The owners of these smartphones, however, do not follow basic security practices. We found that 28% of U.S. smartphone users do not use a screen lock or other safety feature to secure their device. The most commonly used method to secure the device is either a personal identification number (PIN) or a physiological biometric identifier such as a thumbprint [2]. Creating a form of authentication when the smartphone is unlocked is essential given the current utilization rate of smartphone users and the current lack of security practiced by the respective users.

The Department of Defense (DOD) also has a growing need for a form of authentication beyond the common PIN and physiological biometric identifier. Smartphone usage across the DOD, concurrent with the above statistics of adults using smartphones throughout the United States of America, is also on the rise. One field of current interest is within the medical field. The Battlefield Assisted Trauma Distributed Observation Kit runs on a Samsung Galaxy S series Android smartphone. This technology enables simultaneous monitoring and triaging of multiple patients while in combat [3]. The adoption and implementation of modern technological solutions for military problems has been slow at best. A critical concern at the forefront of smartphones and implementation within the DOD is security.

Special Operations Command (SOCOM) has urgent needs in which smartphone technology can and should be applied. SOCOM, in collaboration with the Defense Advanced Research Projects Agency (DARPA), has begun to address the opportunities in which smartphone technology can be applied to SOCOM's mission set. At the forefront of this initiative, DARPA is researching and developing applications specifically for a SOCOM operator. The TransApp program is at the tip of the spear in this union of technology and tactical application [4]. TransApp features many embedded applications;

however, just as the overarching issue with security is a concern holistically for the DOD, it too is a concern for SOCOM. Security and authentication is a concern when the information contained within the smartphone or smart device directly places lives at risk.

The requirement or methodology for unconventional methods of communication is also becoming paramount in times in which communication outages are experienced. Bypassing an already congested and contested electromagnetic spectrum, the utilization of smartphone data as it is interpreted could yield a covert channel of communication. Collecting the measurements and reading the data generated from the sensors could produce a communication medium in the absence of cellular connectivity and coverage.

B. THESIS OBJECTIVES AND APPROACH

The objective of this thesis is to examine the viability of authenticating a known user through the behavioral biometric signature of how each subject orients and holds a smartphone in order to read a headline article. To achieve this objective, this thesis seeks to identify the existence of a unique behavioral biometrics signature when different users interact with a smartphone from the same, fixed location. The interaction is limited to the user's approach in how they handle a smartphone in order to read what is displayed on the screen. The hypothesis is that each user uniquely interfaces with the smartphone based on a confluence of factors to include, but not limited to, height, vision, torso, arm length, and posture. This thesis does not detail each user's unique physical characteristics. This detail is recommended as future work within Chapter V.

C. RELATED WORK

Research within the field of smartphone sensors has primarily revolved around the concept of physical or emotional activity recognition unique to a known user. These works have included an in-depth analysis to determine the physical or emotional state of the known user. These physical states have included—but, not limited to— authenticating a user from their unique walking gait [5], activity and movement recognition [6], and multi-sensor authentication for smartphone security based on machine learning algorithms [7].

In the multi-sensor authentication, the authors Lee and Lee describe in detail the various performance factors of multi-sensor authentication [7]. Their work builds upon the notion of using biometric behavior in order to authenticate a user. They utilize a Nexus 5 Android's orientation, magnetometer, and accelerometer to distinguish a known user.

Lee and Lee collectively used the sensors embedded on the smartphone in ways that tested behavioral markers. This was done through the application of multiple sensors or requiring the user to walk. Gait recognition requires the user to be in motion and the sensor to actively monitor the data from the accelerometer.

In 2016, Alzubaidi and Kalita surveyed existing research on the topic of using behavioral biometrics to authenticate users [8]. Their deep dive of authentication using behavioral biometrics of smartphone users profiled seven types of behavioral biometrics that had been researched to date, including: hand waving, gait, touch-screen, keystroke, voice, signature and general profiling. Amongst their findings were different types of profiling methods. The one profiling method which is noticeably absent is how individuals orient and position their smartphone in order to read text. This feature is the most profound feature utilized in a smartphone today. Smartphones are used less as a phone and more for their common data applications such as Facebook, Instagram, or Twitter.

D. THESIS ORGANIZATION

The remainder of this thesis is organized as follows. In Chapter II, the integration of sensors within smartphones, and the associated trends towards biometric authentication are introduced. The proposed scheme for determining a known subject is described in detail in Chapter III, and the results of the thesis are detailed in Chapter IV. Finally, the thesis is concluded in Chapter V, where significant results and recommendations for future work are presented.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

A. SMARTPHONES

Smartphones have become ubiquitous and woven into the fabric of modern society as part of our everyday lives. Smartphones have created a connection between the user and others through means that are unique to this generation. The personally identifiable information, once aggregated, paints an entire picture of the individual within one device. The smartphone represents a gateway to the individual's financial standing, social media presence, and Internet web browsing history. These devices collect an enormous amount of data on the individual user through sensors embedded in the smartphone.

1. Sensors

The advancement in technology of sensors that are embedded within smartphones has enhanced the user experience of smartphones. A detailed listing of the sensors within the Samsung Galaxy S Series, shown in Table 1, illustrates this evolution.

As smartphones have increased in usage, the sensors have increased the precision of readings to provide an intimate experience for the user. Of note, the S4 is the lone device in the S Series that offers sensors that measure environmental factors (thermometer and humidity). Devices from the S5 forward pivoted toward sensors that related to the user and provided biometric feedback. These sensors—heart rate and fingerprint—introduced physical biometric feedback from the user. The fingerprint, as discussed previously in Chapter I, is a primary feature in the authentication of a user to unlock the device. The recently released Galaxy S8 features technology which leverages the camera sensor in order to authenticate the user via retina scanning and verification. These improvements and enhancements are enabling the advancement of augmented reality that are driving innovation in the user experience with the smartphone.

Table 1. Samsung Galaxy S Series Sensor Evolution. Adapted from [9]

Samsung Galaxy S Series	S1	S2	S3	S4	S5	S6	S7
Accelerometer	✓	✓	✓	✓	✓	✓	✓
GPS	✓	✓	✓	✓	✓	✓	✓
Magnetometer	✓	✓	✓	✓	✓	✓	✓
Light Meter	✓	✓	✓	✓	✓	✓	✓
Proximity	✓	✓	✓	✓	✓	✓	✓
Battery Temperature	✓	✓	✓	✓	✓	✓	✓
Touchscreen	✓	✓	✓	✓	✓	✓	✓
Camera	✓	✓	✓	✓	✓	✓	✓
Cellular Radio	✓	✓	✓	✓	✓	✓	✓
Wifi Radio	✓	✓	✓	✓	✓	✓	✓
Bluetooth	✓	✓	✓	✓	✓	✓	✓
Gyroscope		✓	✓	✓	✓	✓	✓
NFC			✓	✓	✓	✓	✓
Barometer			✓	✓	✓	✓	✓
Pedometer				✓	✓	✓	✓
Thermometer				✓			
Humidity				✓			
Gesture				✓	✓	✓	✓
Color Meter					✓	✓	✓
Heart Rate					✓	✓	✓
Fingerprint					✓	✓	✓
Oxygen Saturation						✓	✓
Magnetic Secure Transmission						✓	✓

2. ACCELEROMETER

One of the sensors most instrumental in the user's smartphone experience is the accelerometer. Most commonly understood as the sensor that determines the orientation of the smartphone as held by the user, the accelerometer serves a critical role within the suite of sensors. The accelerometer's functionality has grown with the emergence of wearable technology, too. A Common application of the accelerometer is gait recognition which is most commonly used with fitness applications that record the user's daily steps.

The accelerometer is a piece of hardware found within the cell phone. In this thesis, the Samsung Galaxy S7 is utilized and the accelerometer chip associated with the S7 is the STMicroelectronics LSM6DS3 [10]. Key features of the LSM6DS3 include a

power consumption of 0.9 mA in combo normal mode and 1.25 mA in combo high-performance mode. The accelerometer is always on due to the lower consumption of power. Embedded within the 2.5 mm x 3 mm x 0.83 mm chip is a temperature sensor [11].

The acceleration chip measures acceleration to include both physical acceleration and gravity. The measurement is reported in the X, Y, and Z fields. These fields most closely resemble pitch, roll, and yaw, which are commonly referenced in the aviation community. All values are in International System of Units (SI) (m/s^2) and measure the acceleration of the device minus the force of gravity along the 3 sensor axes [12].

Using the standard coordinate system, shown in Figure 1, the following conditions apply when the device is at rest on a flat table:

- If the device is pushed to the right, the X acceleration value is positive,
- If the device is pushed away from the user, the Y acceleration value is positive, and
- If the device is lifted off the table toward the user, the Z acceleration value is positive.

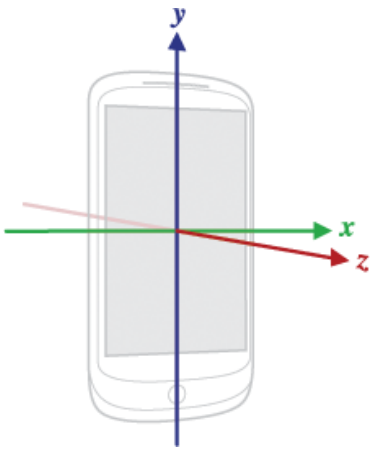

Figure 1. Coordinate System Relative to Mobile Device Used by Accelerometer

Conceptually, an acceleration sensor determines the acceleration that is applied to a device A_D by measuring the forces F_s that are applied to the accelerometer using the relationship [14]

$$A_D = -\left(\frac{1}{mass}\right)\sum F_s \tag{1}$$

where *mass* is the weight of the smartphone. The mass of the Samsung Galaxy S7 is 5.36 ounces [13].

However, the force of gravity g always influences the measured acceleration according to the following relationship:

$$A_D = -g - \left(\frac{1}{mass}\right)\sum F_s \, . \tag{2}$$

For this reason, when the device is at rest on a table, the accelerometer reads approximately 0.0 m/s^2 along the X axis, 0.0 m/s^2 along the Y axis, and 9.8 m/s^2 along the Z axis due to the magnitude of g = 9.81 m/s^2 [14]. In this thesis, the relationship between the force of gravity relative to the orientation of the device as held by the individual user is of most value and is measured and recorded. This is due to the fact that once users interact with the device, they are requested to remain as stationary as possible to record the position and orientation in which individuals most feel comfortable eliminating any other contributions to F_s.

B. AUTHENTICATION

Authentication in the context of security, takes into account three primary strategies. These include knowledge-based, possession for object-based, and biometric. Knowledge-based authentication uses something that is known to the individual; this most commonly is a PIN. Possession for object-based authentication commonly uses a token or ID card. Finally, the biometric strategy is a physical or behavioral characteristic unique to the individual [8]. Examples of both physical and behavioral biometrics are listed in Figure 2. This thesis will explore biometric authentication with a focus on behavioral biometrics.

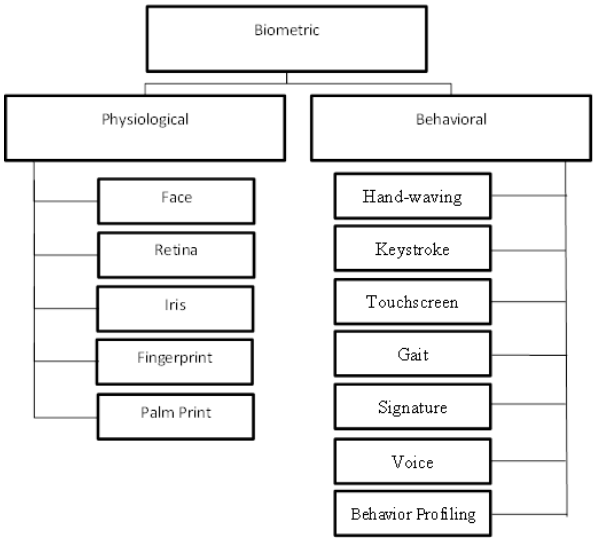

Figure 2. Biometric Authentication Methods. Adapted from [8].

1. Behavioral Biometrics

Behavioral biometrics provide the opportunity for continuous authentication beyond the initial authentication. Behavioral biometrics utilize behavioral traits of a subject. Each subject profiled is anticipated to differ from other subjects when utilizing one or more of the features as shown in Figure 2. Previous work, as discussed in Chapter I, explored these behavioral biometrics across different tests through the utilization of various sensors embedded within the smartphone as detailed in Chapter II, Section A.

C. MATLAB MOBILE

This thesis is centered on the data measured by a smartphone accelerometer. Many applications that measure this data are available on Google play; however, this thesis utilizes MATLAB Mobile created by The MathWorks, Inc [15]. The decision for to use MATLAB Mobile centered on the usability of the data once it is imported into MATLAB for analysis, visualization, and simulation. Additionally, the feature to wirelessly connect the smartphone to a computer with MATLAB installed was essential. MATLAB Mobile and computer communicate via the IEEE 802.11 protocol while each are connected to the same local network. This wireless interfacing enables immediate

feedback. The immediate feedback improves accuracy and reliability of the data from each subject, which is essential while conducting field testing.

MATLAB Mobile has the ability to acquire the following sensor data: acceleration on 3–axes, angular velocity on 3-axes, magnetic field on 3-axes, orientation, and position [16]. The most current version, 4.1.1 released on July 17, 2017, was used for this thesis.

D. SAVITSKY-GOLAY FILTER

A critical element of the thesis is the smoothing of the raw data from the accelerometer as recorded by MATLAB Mobile. The Savitsky-Golay (SGO) filter is the filter utilized throughout this thesis.

SGO filters are commonly utilized to "smooth out" a noisy signal whose frequency span is large. In this type of application, SGO smoothing filters perform much better than standard averaging finite impulse response (FIR) filters. Although SGO filters are more effective at preserving the pertinent high frequency components of the signal, they are less successful than standard averaging FIR filters at rejecting noise [17]. The desire to use SGO filter, also known as polynomial smoothing, or least-squares smoothing filter, is further cemented because a SGO filter preserves the high frequency content of the filter. Due to the unknown variability of the subjects, it was desired to use a filter that preserved the data in order to fully distinguish a subject from another subject. The integrity of the data from subject to subject is paramount and is what qualitatively differentiates the subjects.

The SGO filter output g_i is defined as [18]

$$g_i = \sum_{n=-n_L}^{n_R} c_n f_{i+n} \tag{3}$$

where n_L is the number of points used to the left of a data point and n_R is the number used to the right of the data values f_i of its nearby neighbors. The idea of SGO filtering is to find filter coefficients c_n that preserve high moments [18].

For this thesis, a c_n of 1 for all indices n was chosen due to the low intrusion on the data values relative to a higher order polynomial. This makes the filter a normalized filter. Other c_n values were implemented; however, performance was remarkably improved with all coefficient values equal to 1. The results smoothed the raw data enough to properly distinguish the state in which the subject obtained postural equilibrium necessary to read a headline on the smartphone.

E. MATLAB MACHINE LEARNING

Essential to this thesis is the ability to take a known set of input and output data and determine whether or not the known user is interfacing with the device. Simulating the ability to authenticate a user is paramount to this thesis. MATLAB's machine learning applications was an additional reason we selected the MATLAB Mobile application for this thesis.

Machine based learning is separated into two unique categories: supervised and unsupervised learning. The divisions between these categories is shown in Figure 3. The basis of this thesis focuses on supervised learning since we have identified the data associated with the subject interfacing with the smartphone. In this instance, for supervised learning, we consider it similar to when a device initially requests a PIN. The authenticated user inputs the PIN and immediately the user is requested to re-enter the PIN for verification purposes. A similar logic is applied when an authenticated user interacts with a device to properly calibrate and orient a known behavioral signature.

The factors that determine which supervised learning algorithm would be used is prediction speed and memory usage. Of all available algorithms, decision trees are the simplest yet provide fast and reliable results and require the least amount of memory relative to all other supervised learning algorithms [19]. Given these considerations, the decision tree was selected as the supervised learning algorithm for this thesis. A consideration for future work is the processing power onboard a smartphone and its capability to process data sets into a classifier.

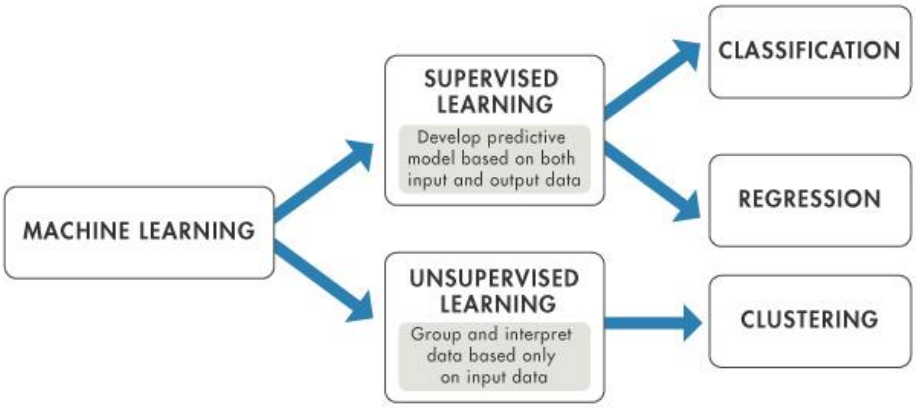

Figure 3. Machine Learning Flow Chart. Source [19].

1. Decision Tree

Decision trees predict responses to data. In order to predict a response, decision trees breakup potential predictions across multiple stages. These stages are represented by root nodes and flow from stage to stage along branches and ultimately arrive at a leaf node, which is the predicted response [20]. An example of a simple decision tree is shown in Figure 4.

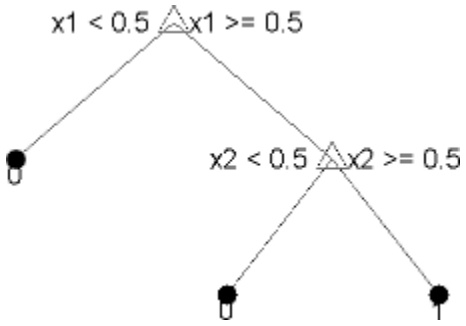

Figure 4. Example of a Simple Classification Tree. Source [20].

In Figure 4, the root nodes are denoted by triangles and the leaf nodes are the corresponding responses—0 and 1. The number of splits (branch points), within the decision tree impact both the complexity and associated predictive response accuracy. As such, the greater number of branches, the longer the time required to train the model.

This thesis examines MATLAB's "Medium Tree" classifier type. A critical feature of the medium tree is the associated number of splits. Relative to a "Simple Tree," a "Medium Tree" has five times the number of splits. Relative to a "Complex Tree," the "Medium Tree" has one fifth of the splits. Given the data set of the thesis, the decision to implement a "Medium Tree" was made [20]. The need for a complex tree was not warranted.

In this thesis, root nodes are represented by the acceleration along the X, Y, and Z axis as recorded by the accelerometer and the leaf node is the associated subject to the predicted response. When the data was compiled and entered into the classifier, leaf node responses were labeled as "Subject 1" or "Not Subject 1."

F. CHAPTER SUMMARY

In this chapter, the technologies instrumental to the thesis were introduced. A discussion of previous generation of smartphone sensors was introduced. As smartphone technology improved so did the sensors within the smartphone. Behavioral biometrics was introduced to understand the difference between a known, common biometric signature and how a behavioral biometric provides value to this thesis. Technological considerations for the thesis were introduced with two features from MATLAB—the mobile application and machine learning application on the desktop computer. These two features make visualization of data (discussed in Chapter IV) seamless. Finally, the SGO filter is introduced and its use as a smoothing filter in order to maximize the unique data obtained from each subject was discussed.

THIS PAGE LEFT INTENTIONALLY BLANK

III. AUTHENTICATING A KNOWN USER USING BEHAVIORAL BIOMETRICS FROM A SMARTPHONE ACCELEROMETER

A. PROPOSED SCHEME

The proposed schemed is one that relies heavily upon repeatability and maximizing controlled variables. The hypothesis began with a simple question, "We (the user) can identify our smartphone, but how can our smartphone identify the user?" While at a local coffee shop, we began to notice that users each oriented the smartphone in a unique manner in order to read the text on the screen. Recall from Chapter II, the orientation of the smartphone can be represented by the X, Y, and Z components of the force of gravity. At a fixed, known location, does a discernable pattern emerge from each user, and if so, what does that pattern resemble?

B. DATA COLLECTION

The basis of the proposed set-up is contingent upon controlling as many variables as possible. Controlling these variables initially reduces the opportunity for complexity. Advanced features and nuances will be addressed in Chapter V and present an opportunity to scale the work.

Prior to the collection of any data, the Human Research Protection Program Office & Institutional Review Board (IRB) at the Naval Postgraduate School (NPS) conducted an official determination and concluded the activity does not involve the use of human subjects because this thesis is not designed to collect information about living individuals. IRB review and Naval Postgraduate School President approval was not required.

At the Starbucks on campus at NPS in Monterey, California, a field testing station was set-up. Each subject was requested to stand in front of the table—shown in Figure 5—and pick up the smartphone. Once the smartphone was picked up, the subject was to then read the headline as shown in Figure 6. This headline article served as a placeholder for a "push notification." Push notifications are very common and are utilized by application software developers to alert the user of an update.

Figure 5. Setup at Starbucks Located at Naval Postgraduate School

Figure 6. Screenshot.of Headline Article from Fox News for Push Notification Field Test

Once the subject read the headline, the subject was requested to retain the position and remain stationary throughout the duration of the data collection. This requirement was critical in order to receive an accurate depiction that the subject was indeed holding the smartphone and to ensure that the data collected while the individual was holding the smartphone had limited potential variance. Due to the desire to only collect data from a position that was fixed, the requirement of additional sensor data was deemed unnecessary.

In the background of the headline article, the MATLAB Mobile application was operational and recording the acceleration data. An example of the MATLAB Mobile application during data collection is displayed in Figure 7.

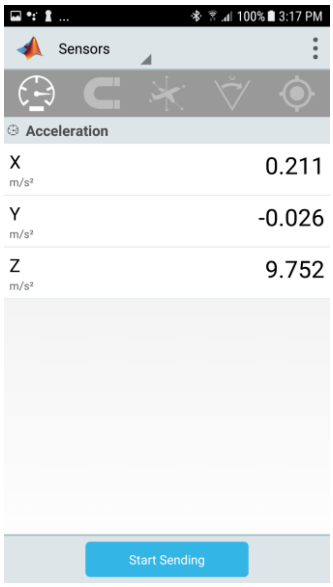

Figure 7. MATLAB Application Displaying Accelerometer Data

The MATLAB code required to establish a connection between the smartphone and workstation is detailed in the Appendix. While receiving the data, the data is logged and converted into a matrix. Figure 8 is a visual representation of the data recorded from the accelerometer on the X, Y, and Z axis.

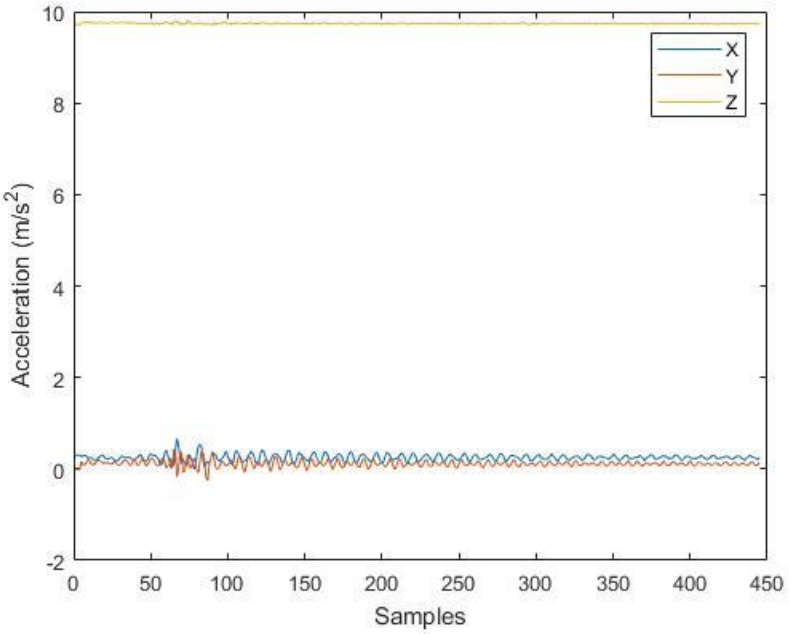

Figure 8. Smartphone Accelerometer Data at Rest

Once the connection between the Galaxy S7 and the computer was confirmed, Subject 1 proceeded to pick up the smartphone from the table shown in Figure 5 and read the headline in Figure 6. Once read, Subject 1 held the smartphone in place and kept it stationary. This behavior is shown in Figure 9. The initial position is the smartphone at rest and is followed by large fluctuations of values in the X, Y, and Z-axes as the smartphone is picked up, and finally stable data while the smartphone was held in a stationary, equilibrium position.

At the recommendation of Android on their developer page, a smoothing filter is applied to the original data set from Figure 9. As discussed in Chapter II, a SGO filter was utilized in the smoothing of the data collected as shown in Figure 10. An opportunity to utilize a different smoothing filter is discussed further in Chapter V and is considered for future work. Additionally, the opportunity to introduce more noise to the signal is also a potential future work.

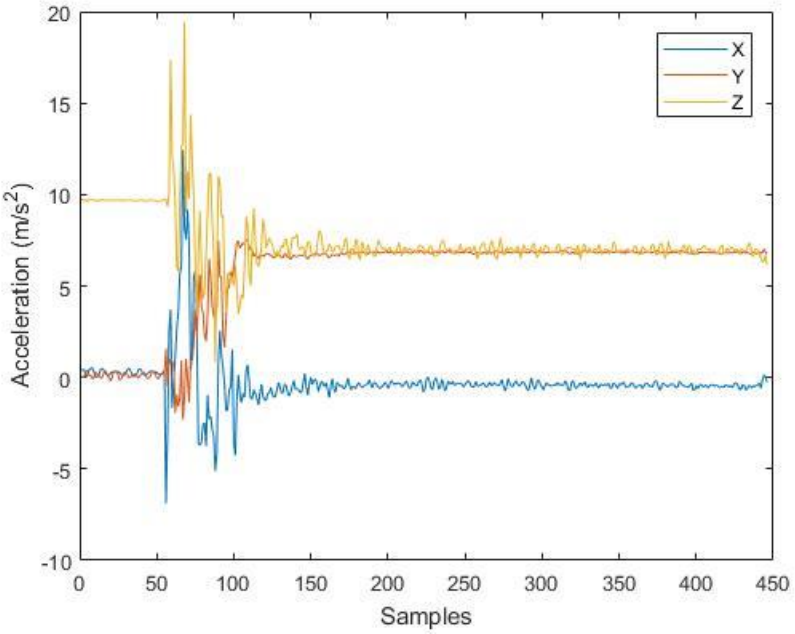

Figure 9. Subject 1 Picking Up Smartphone and Reading Headline Article

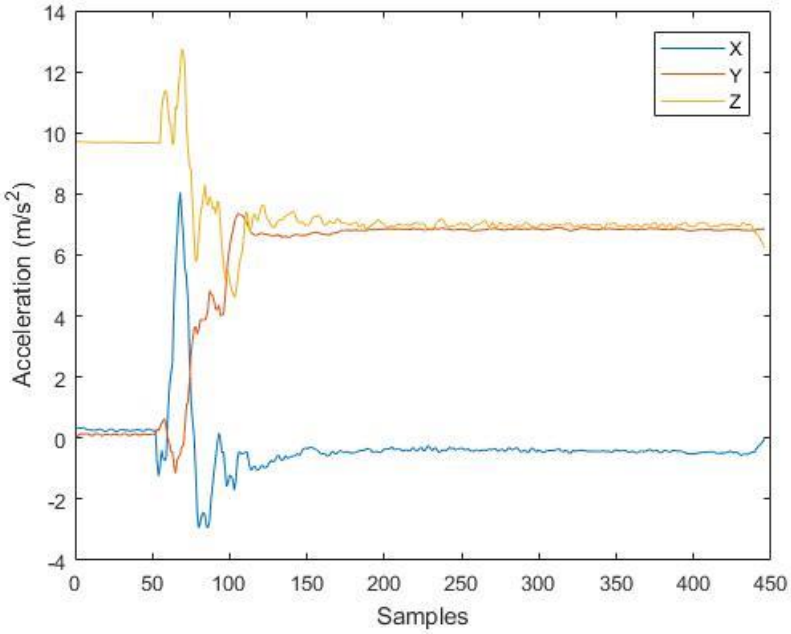

Figure 10. Application of SGO Filter to Figure 9

Once the SGO filter was applied to the data, the data was clipped and 100 samples were chosen from when the smartphone was at equilibrium. 100 samples of Subject 1 holding the phone at equilibrium as selected from Figure 10 is shown in Figure 11.

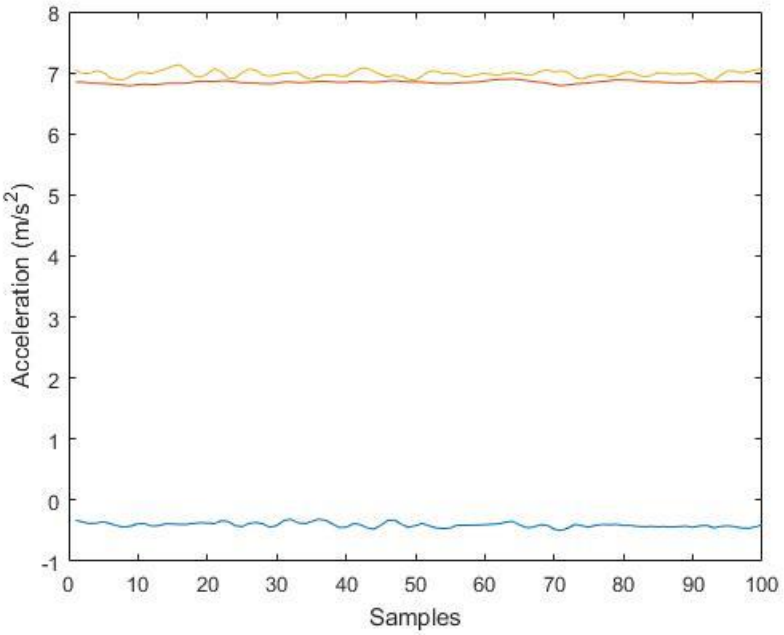

Figure 11. Clipped Data from Subject 1 Holding the Smartphone at Equilibrium

C. CHAPTER SUMMARY

This chapter discussed the method in which data was collected in this thesis in order to determine a known user using behavioral biometrics from a smartphone accelerometer. A SGO FIR filter was applied to smooth the data and a subset of the data at equilibrium was selected for authentication purposes.

IV. ANALYSIS AND RESULTS

A. TESTING

The process from the proposed setup was repeated across 21 different subjects. Subject 1, the authorized user, performed the test for 10 trials. Each trial yielded 100 samples in a similar manner as shown in Figure 11. Subject 2 through Subject 20 each performed 2 trials. These trials were repeated in the same process as Subject 1. Each trial for each Subject yielded 100 samples. Subject 21, an unknown user, also performed 2 trials of interacting with the smartphone. Each trial yielded 100 samples of data, too.

B. RESULTS

Initial examination of the data demonstrated the potential for unique signatures. A comparison of collected data is shown in Figures 12, 13, and 14 across the X, Y and Z axis, respectively. These figures compare Subject 1 and Subject 2's 100 samples in a histogram. The histogram presents the bin count of the occurrences of a measured acceleration value on the associated axis.

Of particular note and interest is the difference between the distribution of Subject 1 and Subject 2 along the Y and Z axis. There is a clear and distinct delta between Subject 1 and 2 in the Y-axis with a less distinguishing difference in the Z, and nearly identical values in the X axis.

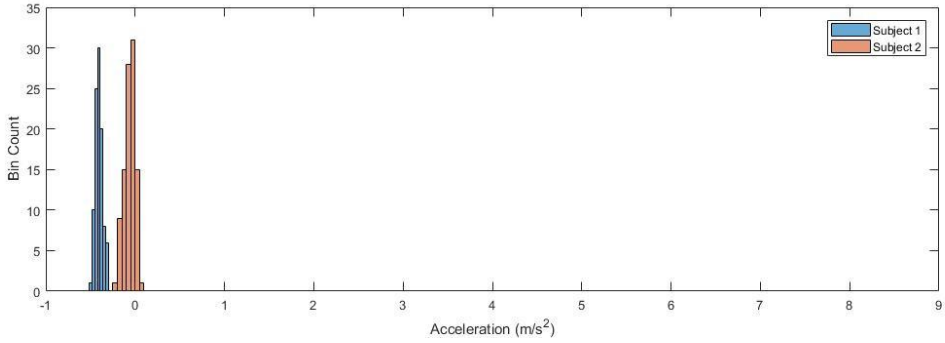

Figure 12. Subject 1 vs. Subject 2 X-Axis Histogram

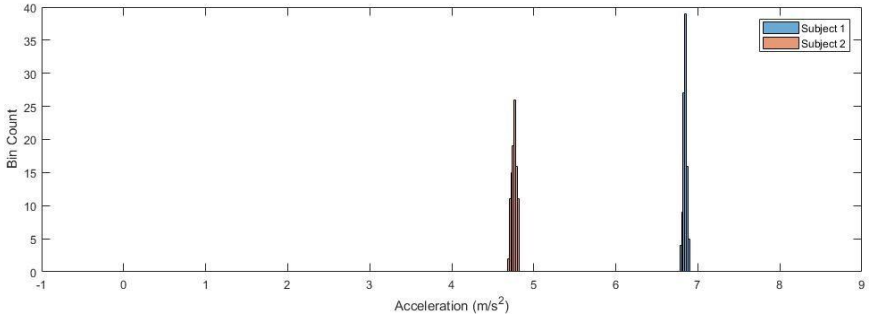

Figure 13. Subject 1 vs. Subject 2 Y-Axis Histogram

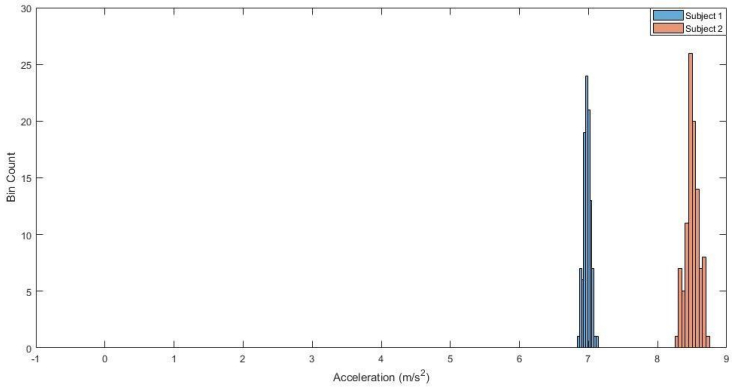

Figure 14. Subject 1 vs. Subject 2 Z-Axis Histogram

While one subject when compared and evaluated with another is easily discernable, once multiple subjects are aggregated, there was less discernable traits. Utilizing two trials and 100 samples from each trial, Subject 1's comparison to the other 19 subjects is shown in the accelerometer histograms in Figures 15, 16, and 17 across the X, Y, and Z axis, respectively.

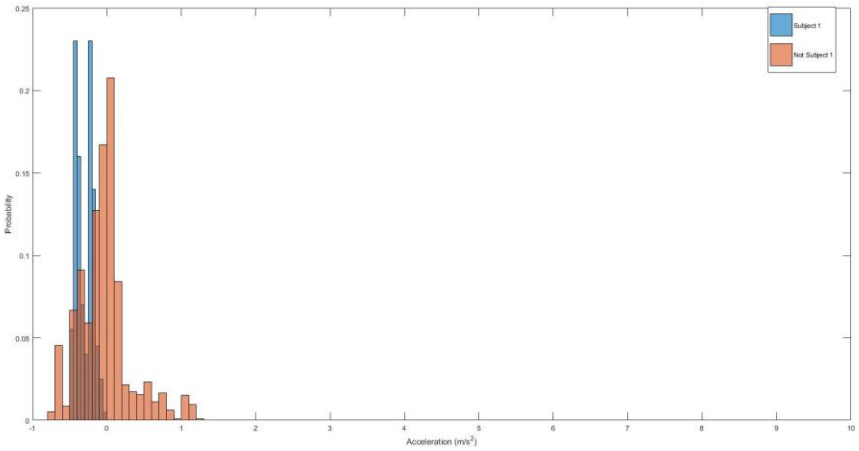

Figure 15. Subject 1 vs. Not Subject 1 X-Axis

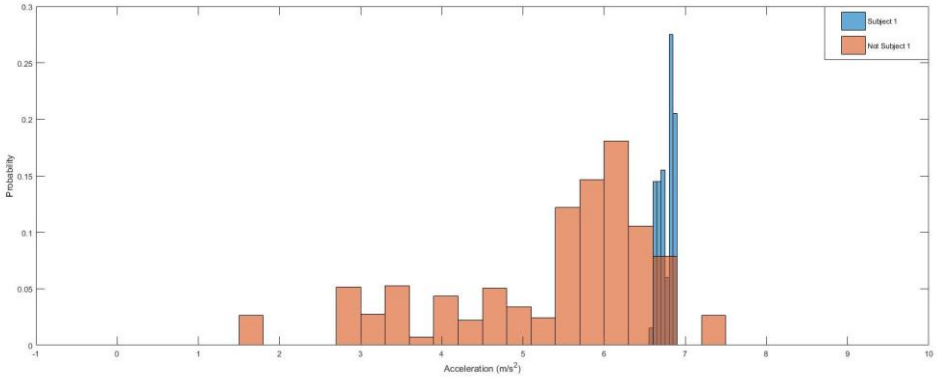

Figure 16. Subject 1 vs. Not Subject 1 Y-Axis

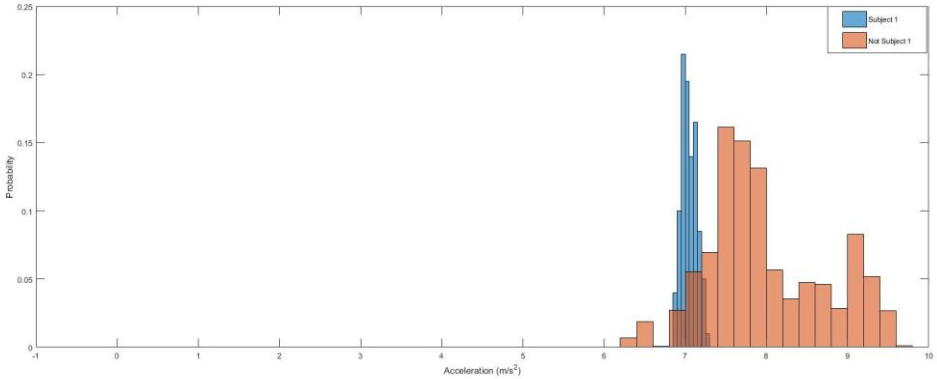

Figure 17. Subject 1 vs. Not Subject 1 Z-Axis

The randomness of the data across the sample of data does not present any discernable pattern of Subject 1 vs. Not Subject 1. Consequently, an unsupervised learning classifier using clustering would be inappropriate in this situation. This absence of a discernable pattern is the motivation behind utilizing MATLAB's machine learning application software. Through supervised learning we are able to properly identify the likelihood the authenticated user, Subject 1, is interfacing with the smartphone.

Across each of the Subjects measured—1 through 20—the greatest return of investment with distinguishing a unique user is on the Y and Z axis. These values are shown in Figure 18. Figure 18 is a scatter plot that plots the Y and Z axis values of 1,000 samples of Subject 1 against 1,900 samples—19 Subjects with 100 samples each—in a visualization.

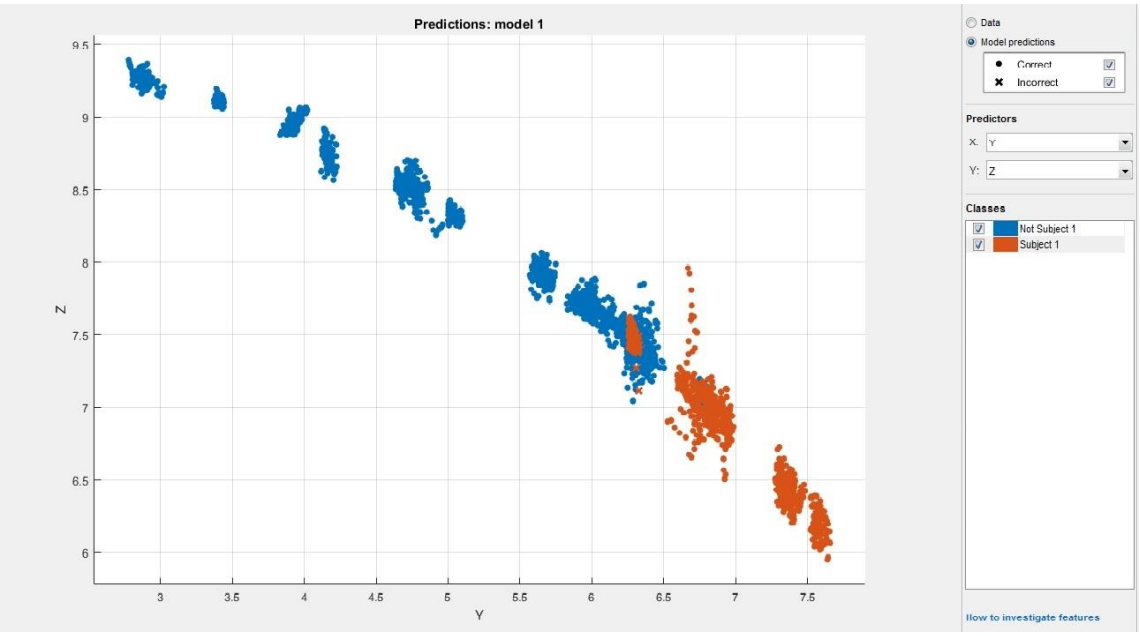

Figure 18. YZ Scatter Plot—Subject 1 vs. Not Subject 1

These samples were injected into a machine learning classification simulation within MATLAB. Of the 2,900 samples, predictions were made of the classifier and the trained model yielded a confidence factor of 98% as shown in Figure 19.

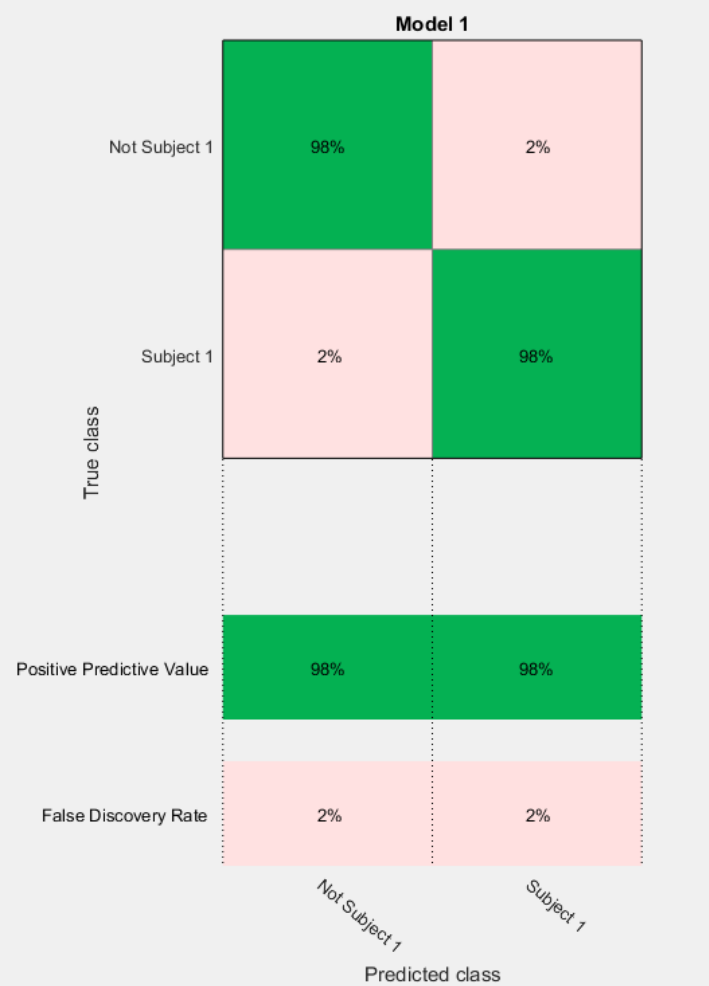

Figure 19. Confusion Matrix of Trained Model

After a model was trained with the 2,900 samples of Subject 1 and Not-Subject 1, 100 samples from trial number 1 of Subject 21 was tested against the classifier. The data, when comparing the Y and Z axis, is very distinguishable between each subject. From Figure 20, the dramatic difference in Y and Z axis data is shown.

Once the 100 samples were entered into the trained model, each sample accurately reported each sample was not Subject 1, the authenticated user.

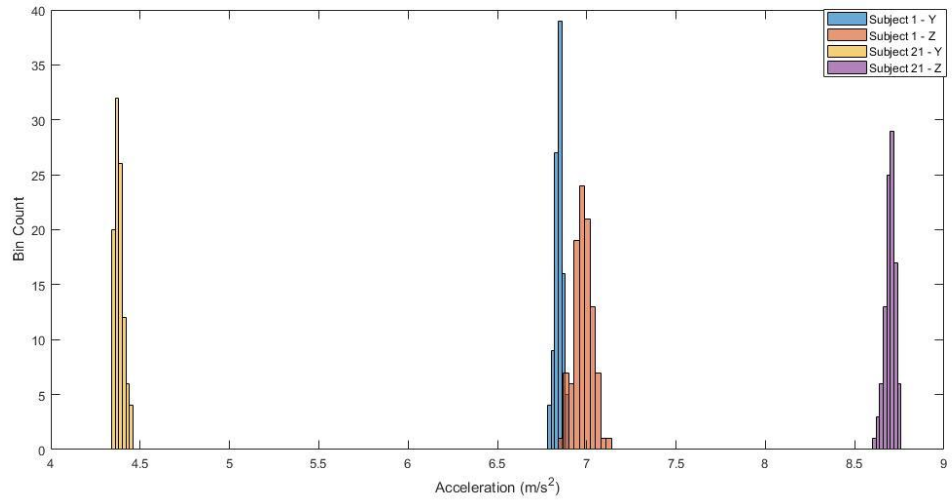

Figure 20. Subject 1 vs. Subject 21 Y & Z Axis

The significance is that with a random pool of individuals selected, there is 98% confidence that a distinction can be made between one known individual and 19 unknown individuals. Utilizing a trained model from the 98% confidence, a new, unidentified user, Subject 21, was then accurately identified as not Subject 1 – the known subject.

C. CHAPTER SUMMARY

In this chapter, an analysis of the data collected in Chapter III is presented. Across the three axes—X, Y, and Z—a pattern is discernable from one subject to another. However, when multiple subjects interface with the smartphone, more uncertainty is introduced. In order to accurately determine the authentication of the smartphone user, a machine learning algorithm—a decision tree—was utilized. This decision tree algorithm produced a 98.0% confidence factor that the authenticated user was accurately predicted by the trained model. Once the model was trained utilizing 20 different subjects, an unknown subject, Subject 21, was introduced to the trained model. As shown in Figure 20, the difference between Subject 1 and Subject 21 is discernable. The trained model accurately validated Subject 21 as not Subject 1 for the entire trial of 100 samples.

V. CONCLUSION AND RECOMMENDATIONS

We have identified a means in which controlling certain variables to identify a user through an uncommon method is possible. Identifying the individual, and perhaps more importantly, knowing that the user interacting with the smartphone is not the intended or authenticated individual is of equal to if not more importance or significance. In examining the profile of individuals who physically picked up a smartphone, oriented the smartphone in order to read text, and maintained that posture, a unique profile for a known user against other unknown users was obtained. Exact understanding of what physical attributes contributed most to the variations in Y and Z axis accelerometer components remains unknown.

A. RECOMMENDATIONS AND FUTURE WORK

While smartphones continue to evolve and improve, the opportunity to re-evaluate this thesis will remain. An area of future work includes determining and evaluating the confidence factor of previous generation smartphone technology relative to current generation technology. This thesis was completed exclusively with a Samsung Galaxy S7. Future work would include comparison of this technology with previous generations of the Galaxy family. Of note, this thesis attempted to utilize the Galaxy S3. The MATLAB application had compatibility issues and presented interface challenges.

This thesis placed a premium on the smoothing of the data utilizing the SGO filter design. Introducing noise into the original signal could distort the results. Evaluating the confidence factor once noise is introduced is another potential future work, As noted during Chapter III, replicating the thesis with a different smoothing filter may yield different results.

Additional experimentation utilizing inputs from other sensors may increase the fidelity of the authentication. Given the parameters of the thesis, a premium was on the user's interaction in a stationary posture while reading a known text script. Further tests are needed to understand how the user interacts with the smartphone beyond reading the

smartphone in a fixed position. Measuring the interaction between users and various applications could be evaluated.

Given the ongoing interaction of the smartphone, continuous authentication and monitoring is possible. The closer we understand the metrics measured that distinguish user from user, not only will security be heightened but, most importantly, the user experience will increase, as well.

APPENDIX. MATLAB CODE FOR EXTRACTING DATA FROM ACCELEROMETER

```
%%Establish new connection with required password
connector off;
connector on password;
m=mobiledev;

%%Discard previous logs
discardlogs(m);

%%Establish sample rate
m.SampleRate=50;

%%Enable sensors
m.AccelerationSensorEnabled = 1;

%%Begin to log, pause, and stop logging the data
m.logging = 1;
pause(15);
m.logging = 0;

%%Disable sensors
m.AccelerationSensorEnabled = 0;

%%Extract logged sensor data
[subject1_1,t]=accellog(m);

%%Apply smoothing filter to extracted sensor data
sgo_subject1_1 = sgolayfilt(subject1_1,1,7);
```

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

[1] J. Poushter, ”Smartphones are common in advanced economies, but digital divides remain.” Pew Research Center, April 214, 2017. [Online]. Available: http://www.pewresearch.org/fact-tank/2017/04/21/smartphones-are-common-in-advanced-economies-but-digital-divides-remain/

[2] L. Rainnie and A. Perrin, (2017, June 28). 10 facts about smartphones as the iPhone turns 10. [Online]. Available: http://www.pewresearch.org/fact-tank/2017/06/28/10-facts-about-smartphones/.

[3] G. Slabodkin. (2017, August 14). Battlefield software supports medics’ efforts to provide care. [Online]. Available: https://www.healthdatamanagement.com/news/battlefield-software-supports-medics-with-care-documentation.

[4] A. Estes. (August 14, 2017). Inside the Military’s Secretive Smartphone Program. [Online]. Available: http://gizmodo.com/inside-the-militarys-secretive-smartphone-program-1603143142.

[5] N. Yodpijit and N. Tavichaiyuth and M. Jongprasithporn, “The use of smartphone for gait analysis,” in *2017 3rd Int.l Conf. on Control, Automation and Robotics*, 2017, pp. 543–546.

[6] M. Jongprasithporn, N. Yodpijit, and R. Srivilai, “A smartphone-based real-time simple activity recognition,” in *2017 3rd International Conference on Control, Automation and Robotics (ICCAR)*, 2017, pp. 539–542.

[7] W. H. Lee and R. B. Lee, “Multi-sensor authentication to improve smartphone security,” in 2015 International Conference on Information Systems Security and Privacy (ICISSP), 2015, pp. 1–11.

[8] A. Alzubaidi and J. Kalita, “Authentication of smartphone users using behavioral biometrics,” *IEEE Communications Surveys & Tutorials*, vol. 18, (3), pp. 1998–2026, 2016.

[9] K. Fitchard. (2016, February 19). Sensing Samsung: The evolution of sensors in the Galaxy S series. [Online]. Available: https://opensignal.com/blog/2016/02/19/sensing-samsung-the-evolution-of-sensors-in-the-galaxy-s-series/

[10] ChipWorks. (2017, September 13). Galaxy S7: Snapdragon Chipsets vs Exynos Chipsets. [Online]. Available: http://www.chipworks.com/about-chipworks/overview/blog/galaxy-s7-snapdragon-chipsets-vs-exynos-chipsets

[11] St. (2017, September 13). LSM6DS3. [Online]. Available: http://www.st.com/en/mems-and-sensors/lsm6ds3.html

[12] Android. (2017, March 27). Sensor types. [Online]. Available: https://source.android.com/devices/sensors/sensor-types

[13] CNet. (2017, September 13). Samsung Galaxy S7 Specs. [Online]. Available: https://www.cnet.com/products/samsung-galaxy-s7/specs/

[14] Android. Motion Sensors. [Online]. Available: https://developer.android.com/guide/topics/sensors/sensors_motion.html

[15] Google Play. (2017, September 13) MATLAB Mobile. [Online]. Available: https://play.google.com/store/apps/details?id=com.mathworks.matlabmobile&hl=en

[16] Math Works. (2017, September 13). MATLAB Mobile. [Online]. Available: https://www.mathworks.com/products/matlab-mobile.html

[17] W. Gander and J. Hrebicek. "Smoothing Filters," in *Solving Problems in Scientific Computing Using Maple and MATLAB*, 3rd ed. Berlin, Germany: Spring-Verlag Berlin Heidelberg, 1997, pp. 135–154.

[18] W. Press, B. Flannery, and S. Teukolsky, *Numerical Recipes in C: the Art of Scientific Computing*, 2nd ed. New York, NY: Cambridge University Press, 1992, pp. 650.

[19] MathWorks. (2017, September 14). What Is Machine Learning? [Online]. Available: https://www.mathworks.com/discovery/machine-learning.html

[20] MathWorks. (2017, September 14). Choose Classifier Options. [Online]. Available: https://www.mathworks.com/help/stats/choose-a-classifier.html

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California