

5-2018

# Evaluation of Pedometer Performance Across Multiple Gait Types Using Video for Ground Truth

Ryan Stephen Mattfeld  
*Clemson University*, [rmattfe@g.clemson.edu](mailto:rmattfe@g.clemson.edu)

Follow this and additional works at: [https://tigerprints.clemson.edu/all\\_dissertations](https://tigerprints.clemson.edu/all_dissertations)

---

## Recommended Citation

Mattfeld, Ryan Stephen, "Evaluation of Pedometer Performance Across Multiple Gait Types Using Video for Ground Truth" (2018).  
*All Dissertations*. 2122.  
[https://tigerprints.clemson.edu/all\\_dissertations/2122](https://tigerprints.clemson.edu/all_dissertations/2122)

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact [kokeefe@clemson.edu](mailto:kokeefe@clemson.edu).

# EVALUATION OF PEDOMETER PERFORMANCE ACROSS MULTIPLE GAIT TYPES USING VIDEO FOR GROUND TRUTH

---

A Dissertation  
Presented to  
the Graduate School of  
Clemson University

---

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy  
Electrical Engineering

---

by  
Ryan Stephen Mattfeld  
May 2018

---

Accepted by:  
Dr. Adam W. Hoover, Chair  
Dr. Elliot Jesch  
Dr. Richard E. Groff  
Dr. Ian Walker

# Abstract

This dissertation is motivated by improving healthcare through the development of wearable sensors. This work seeks improvement in the evaluation and development of pedometer algorithms, and is composed of two chapters describing the collection of the dataset and describing the implementation and evaluation of three previously developed pedometer algorithms on the dataset collected. Our goal is to analyze pedometer algorithms under more natural conditions that occur during daily living where gaits are frequently changing or remain regular for only brief periods of time. We video recorded 30 participants performing 3 activities: walking around a track, walking through a building, and moving around a room. The ground truth time of each step was manually marked in the accelerometer signals through video observation. Collectively 60,853 steps were recorded and annotated. A subclass of steps called shifts were identified as those occurring at the beginning and end of regular strides, during gait changes, and during pivots changing the direction of motion. While shifts comprised only .03% of steps in the regular stride activity, they comprised 10-25% of steps in the semi-regular and unstructured activities. We believe these motions should be identified separately, as they provide different accelerometer signals, and likely result in different amounts of energy expenditure. This dataset will be the first to specifically allow for pedometer algorithms to be evaluated on unstructured gaits that more closely model natural activities.

In order to provide pilot evaluation data, a commercial pedometer, the Fitbit Charge 2, and three prior step detection algorithms were analyzed. The Fitbit consistently underestimated the total number of steps taken across each gait type. Because the Fitbit algorithm is proprietary, it could not be reimplemented and examined beyond a raw step count comparison. Three previously published step detection algorithms, however, were implemented and examined in detail on the dataset. The three algorithms are based on three different methods of step detection; peak detection, zero crossing (threshold based), and autocorrelation. The evaluation of these algorithms was performed across 5

dimensions, including algorithm, parameter set, gait type, sensor position, and evaluation metric, which yielded 108 individual measures of accuracy. Accuracy across each of the 5 dimensions were examined individually in order to determine trends. In general, training parameters to this dataset caused a significant accuracy improvement. The most accurate algorithm was dependent on gait type, sensor position, and evaluation metric, indicating no clear “best approach” to step detection. In general, algorithms were most accurate for regular gait and least accurate for unstructured gait. In general, accuracy was higher for hip and ankle worn sensors than it was for wrist worn sensors. Finally, evaluation across running count accuracy (RCA) and step detection accuracy (SDA) revealed similar trends across gait type and sensor position, but each metric indicated a different algorithm with the highest overall accuracy.

A classifier was developed to identify gait type in an effort to use this information to improve pedometer accuracy. The classifier’s features are based on the Fast Fourier Transform (FFT) applied to the accelerometer data gathered from each sensor throughout each activity. A peak detector was developed to identify the maximum value of the FFT, the width of the peak yielding the maximum value, and the number of peaks in each FFT. These features were then applied to a Naive Bayes classifier, which correctly identified the gait (regular, semi-regular, or unstructured) with 84% accuracy. A varying algorithm pedometer was then developed which switched between the peak detection, threshold crossing, and autocorrelation based algorithms depending on which algorithm performed best for the sensor location and detected gait type. This process yielded a step detection accuracy of 84%. This was a 3% improvement when compared to the greatest accuracy achieved by the best performing algorithm, the peak detection algorithm. It was also identified that in order to provide quicker real-time transitions between algorithms, the data should be examined in smaller windows. Window sizes of 3, 5, 8, 10, 15, 20, and 30 seconds were tested, and the highest overall accuracy was found for a window size of 5 seconds. These smaller windows of time included behaviors which do not correspond directly with the regular, semi-regular, and unstructured gait activities. Instead, three stride types were identified: steady stride, irregular stride, and idle. These stride types were identified with 82% accuracy. This experiment showed that at an activity level, gait detection can improve pedometer accuracy and indicated that applying the same principles to a smaller window size could allow for more responsive real-time algorithm selection.



# Acknowledgments

I would like to give special thanks my advisor, Dr. Adam Hoover, for the numerous hours of guidance and advice provided from the beginning to the end of my journey through graduate school at Clemson University. He assisted in the development of my technical writing skills, research techniques, and problem solving ability, all of which have been critical in my development as a researcher. I would also like to thank my committee members, Dr. Groff, Dr. Jesch, and Dr. Walker for the time and effort spent reviewing my dissertation. Additional thanks go out to Daniel Noneaker, for helping encourage me to pursue graduate school and supporting me administratively throughout, as well as numerous other professors in the ECE department who set excellent teaching examples and supported my education.

In addition, I would like to thank my wife, Valerie, my parents, Ann and Steve, and my sister, Amanda, for being incredibly supportive throughout my studies. Several colleagues and friends have been instrumental in my endeavors, including Surya Sharma, Raul Ramos-Garcia, and JP Kwon. Several other friends and colleagues have assisted me throughout my graduate studies, including Daniel Cutshall, Shawn Mathew, Jay Puthumanappilly, and Mohan Ramaraj, among many others.

This project was made possible due to the assistance of a grant provided by the Brooks Sports Science Institute at Clemson University.

# Table of Contents

Title Page . . . . .	i
Abstract . . . . .	ii
Acknowledgments . . . . .	iv
Table of Contents . . . . .	v
List of Tables . . . . .	vii
List of Figures . . . . .	ix
<b>1 Introduction . . . . .</b>	<b>1</b>
1.1 Overview . . . . .	1
1.2 MEMS sensors . . . . .	5
1.2.1 Accelerometers . . . . .	5
1.2.2 Gyroscopes . . . . .	9
1.3 Activity Recognition . . . . .	9
1.4 Pedometer Algorithm Evaluation . . . . .	12
1.5 Contribution and Novelty . . . . .	18
<b>2 Data Collection and Evaluation of Steps, Shifts, and Fitbit . . . . .</b>	<b>20</b>
2.1 Methods . . . . .	20
2.1.1 Data Collection . . . . .	21
2.1.2 Patterns in Regular Gait . . . . .	34
2.1.3 Challenges in natural gaits . . . . .	37
2.1.4 Ground Truthing . . . . .	40
2.1.5 Metrics . . . . .	43
2.1.6 Definition of a Step . . . . .	46
2.2 Results . . . . .	47
2.2.1 Steps and Shifts . . . . .	48
2.2.2 Fitbit Accuracy . . . . .	49
2.3 Conclusions and Future Work . . . . .	51
<b>3 Implementation of Prior Pedometer Algorithms . . . . .</b>	<b>54</b>
3.1 Methods . . . . .	54
3.1.1 Step Detection Algorithms . . . . .	54
3.1.2 Implemented Algorithms . . . . .	61
3.1.2.1 Peak Detector . . . . .	61
3.1.2.2 Threshold Based . . . . .	63
3.1.2.3 Autocorrelation . . . . .	65

3.1.3	Parameter Tuning . . . . .	66
3.2	Results . . . . .	70
3.2.1	Evaluation of Parameter Training . . . . .	72
3.2.2	Evaluation of Metrics . . . . .	73
3.2.3	Evaluation of Algorithms . . . . .	76
3.2.4	Evaluation of Gait Types . . . . .	77
3.2.5	Evaluation of Sensor positions . . . . .	78
3.3	Conclusions and Future Work . . . . .	78
<b>4</b>	<b>Improving Pedometer Performance Through Detection of Gait Type . . . . .</b>	<b>81</b>
4.1	Methods . . . . .	81
4.1.1	Fast Fourier Transform . . . . .	82
4.1.2	Gait vs Stride . . . . .	86
4.1.3	Classification . . . . .	90
4.1.4	Varying Pedometer Algorithm Based on Detected Gait . . . . .	91
4.2	Results . . . . .	91
4.2.1	Gait Detection . . . . .	92
4.2.2	Stride Detection . . . . .	94
4.2.3	Varying Pedometer Algorithms . . . . .	99
4.3	Conclusions and Future Work . . . . .	101
<b>5</b>	<b>Conclusion . . . . .</b>	<b>103</b>
	<b>Appendices . . . . .</b>	<b>105</b>
A	Additional Work . . . . .	106
A.1	HMD Latency Analysis . . . . .	106
A.2	Table Embedded Scale . . . . .	107
B	Publications . . . . .	109
C	PAR-Q . . . . .	110
	<b>Bibliography . . . . .</b>	<b>111</b>

# List of Tables

1.1	Summary of evaluations performed for pedometer algorithm development. (Numbers in parentheses indicate variations within the location or pace indicated. For example, Walk(3) indicated that 3 different walking paces were examined and Hand(3) indicates that the sensor was held in the hand in three different orientations or postures) . . .	13
2.1	Summary of steps and shifts manually recorded through each of the three activities.	48
2.2	Comparison of ground truth steps with number of steps reported by Fitbit. . . . .	50
2.3	Comparison of ground truth steps, excluding shifts, with the number of steps reported by Fitbit. . . . .	51
3.1	A summary of prior algorithm methodologies and identification of the algorithms selected for reimplementatation. . . . .	55
3.2	A listing of sample spaces searched for each variable used in the peak detection pedometer algorithm. . . . .	69
3.3	A listing of sample spaces searched for each variable used in the threshold based pedometer algorithm. . . . .	69
3.4	A listing of sample spaces searched for each variable used in the autocorrelation pedometer algorithm. . . . .	69
3.5	Accuracy results across each of the five variables used for evaluating prior algorithms: Parameters (default or trained), evaluation metric (RCA or SDA), gait (regular, semi-regular, or unstructured), sensor position (wrist, hip, or ankle), and algorithm (peak detection, threshold based, or autocorrelation). . . . .	73
3.6	A comparison of the number of time default parameters or tuned parameters provided better results. . . . .	73
3.7	The effect of evaluation metric on determining algorithm performance. . . . .	74
3.8	The effect of evaluation metric on performance according to sensor position . . . . .	74
3.9	The effect of evaluation metric on performance according to gait type. . . . .	75
3.10	The most accurate algorithm (peak detection, theshold, or autocorrelation), according to RCA, for each combination of sensor position and gait type. Results use parameters trained on regular RCA for the wrist. . . . .	76
3.11	The most accurate algorithm (peak detection, theshold, or autocorrelation), according to SDA, for each combination of sensor position and gait type. Results use parameters trained on regular RCA for the wrist. . . . .	77
3.12	The effect of gait type on accuracy. . . . .	78
3.13	The effect of evaluation metric on performance according to sensor position . . . . .	78
4.1	Confusion Matrix for Gait Detection. . . . .	92
4.2	Gait detection accuracy with respect to sensor location. . . . .	93
4.3	Stride detection accuracy using a 5 second window with respect to sensor location. . . . .	95
4.4	Confusion matrix for stride detection. . . . .	96
4.5	The ground truth time for each stride type, in minutes. . . . .	96

4.6	Stride detection accuracy using a 5 second window, categorized by gait type. . . . .	97
4.7	Step detection accuracy while varying the algorithm based upon classifier detection of gait type. When compared to the most accurate individual algorithm within each metric, RCA is the same, and SDA demonstrates a 3% improvement. . . . .	100
4.8	Step detection accuracy while varying algorithm based upon classifier detection of gait type, as compared to the average of three algorithms' performances, across sensor location. All accuracies show large improvements. . . . .	100
4.9	Step detection accuracy while varying algorithm based upon classifier detection of gait type, as compared to the average of three algorithms' performances, across gait type. All accuracies show large improvements. . . . .	101

# List of Figures

1.1	The x, y, z coordinate system used by the Shimmer3 device to track acceleration. . . . .	6
1.2	A mass spring system. . . . .	7
1.3	A pedometer made using the spring mass system to determine when steps occur. . . . .	8
1.4	The roll, pitch, and yaw system used to define the measurement from a gyroscope. . . . .	10
1.5	Sample accelerometer signal typical in regular gait analysis. . . . .	11
2.1	The Shimmer3 device used for motion data collection. . . . .	23
2.2	The base used to connect the Shimmer3 device to the PC and transfer data. . . . .	23
2.3	A participant wearing the three Shimmer3 devices, one each on the wrist, hip, and ankle. . . . .	24
2.4	The procedure used in order to record all steps taken by the participant through each activity. . . . .	25
2.5	The image produced by the recording process. . . . .	26
2.6	The path followed during the regular gait activity is identified with a red line (the portion of the sidewalk hidden by the building is indicated by a dashed line). . . . .	27
2.7	One of the toy footballs that participants searched for during the semi-regular gait activity. . . . .	28
2.8	One of the hallways in Riggs Hall that participants traveled during the semi-regular gait activity. . . . .	29
2.9	One of the rooms searched by participants during the semi-regular gait activity. . . . .	30
2.10	The original Lego instruction booklet, nearly assembled Lego, and bin containing the final pieces for assembly. . . . .	31
2.11	A visualization of the process used throughout the unstructured gait activity. Participants construct Legos in a central location while periodically collecting additional bins of lego pieces located around the room. . . . .	32
2.12	A printed copy of the Lego instruction booklet, with handwritten numbers in the top right of each frame indicating which bin contains the pieces necessary for completing the corresponding step. . . . .	33
2.13	The Fitbit Charge 2 was worn by each participant to assess the accuracy of a commercial device in each activity. . . . .	34
2.14	Example of acceleration measured for each location and axis. Acceleration peaks related to each step are most visible in the Wrist Y axis and Hip Y axis. For the ankle, the characteristic step pattern is a period of no motion (as the non-instrumented step is moving), followed by two periods of acceleration as the instrumented begins moving and stops moving. . . . .	35
2.15	Examples of finding the magnitude of an accelerometer signal divided across a 3-axis accelerometer. . . . .	38
2.16	Example of different acceleration signals for the wrist (top), hip (middle), and ankle (bottom) during regular walking. . . . .	39
2.17	Examples of accelerometer data collected from the three different activities performed. . . . .	40

2.18	The tool developed for visualizing the accelerometer data and identifying ground truth steps. The 9 axes of accelerometer signal (x, y, and z axis for wrist, hip, and ankle) are displayed on the left, and the video recording is displayed in the top right. Steps taken with the right foot are marked with a solid red line and steps taken with the left foot are marked with a dashed blue line. . . . .	42
2.19	An example of ground truth steps identified in the regular gait activity. The peak in acceleration between each right step (red solid line) and left step (blue dashed line) is indicative of the forward motion of the left leg. . . . .	46
2.20	An example of the accelerometer signal associated with four different types of shifts. The first two rows provide examples of first and last steps being taken. The third row demonstrates a pivot. The fourth row demonstrates a shuffle step. . . . .	47
3.1	Example of peaks detected in an accelerometer signal taken from the hip during regular gait. Peaks are indicated with arrows. . . . .	56
3.2	Example of zero crossing detection in an accelerometer signal taken from the hip during regular gait. Red and green dashed lines represent adaptive maximum and minimum based on the previous segment of data. The solid blue lines indicate the mean of the maximum and minimum and are used as the threshold for zero crossing detections. Orange circles indicate when the accelerometer signal crosses the threshold, indicating a step. . . . .	58
3.3	An example of obtaining an maximum autocorrelation window and score for index 1200. Starting at index 1200, a range of window sizes are tested for autocorrelation. The smallest window, indicated by a red solid line, is compared against the following window (indicated by a red dashed line). The correlation between these two windows is recorded and compared against correlation scores from a range of window sizes. The highest correlation score and its corresponding window size are recorded and evaluated in the algorithm. . . . .	59
3.4	Example of how a the foot moves when walking. Gyroscope algorithms are able to measure the rotation in the foot present in this process. . . . .	60
3.5	Flowchart of the peak detection algorithm. Parameters used in the algorithm are listed, with the values used in the publication in bold. . . . .	62
3.6	Flowchart of the threshold based algorithm. Parameters used in the algorithm are listed, with the values used in the publication in bold. . . . .	64
3.7	Flowchart of the autocorrelation algorithm. Parameters used in the algorithm are listed, with the values used in the publication in bold. . . . .	65
3.8	Demonstration of the complexity involved in pedometer evaluation, give the dimensionality of the problem. . . . .	71
3.9	The five variables that can be examined and adjusted when evaluating pedometer algorithms on the proposed dataset. The number in parentheses indicates the number of options examined for each variable. . . . .	71
4.1	An example of an FFT generated by a regular gait activity. . . . .	83
4.2	FFT results for one participant, for each sensor location (wrist, hip, and ankle) and each gait type (regular, semi-regular, and unstructured). FFT calculated using the magnitude of acceleration (including X, Y, and Z axes). . . . .	84
4.3	An example of an FFT generated by a regular gait activity, wrist-worn sensor. This FFT is difficult to distinguish from a semi-regular gait FFT. . . . .	86

4.4	An example of the process of creating a ground truth for stride detection (steady stride, irregular stride, idle). The process is based on the ground truth step indices. There are three thresholds for detection: time between steady stride steps ( $T_{ss}$ ), number of consecutive steps needed for steady stride ( $N_{ss}$ ), and time to transition between irregular steps and idle ( $T_{is}$ ). Each of these thresholds were determined heuristically. . . . .	87
4.5	A demonstration of the window-based ground truth determination process. In the example, a 5 second window is approximated, and the window's ground truth value is determined by a majority vote of individual index ground truth values. In the second to last window shown, the window's ground truth value is very close to either irregular stride or steady stride, demonstrating how, in some cases, small changes in one or two step indices can change the ground truth value for a window. . . . .	88
4.6	Example of 1 minute of motion in a semi-regular gait activity. Motion is broken in to 5 seconds windows and each window of activity is converted to the frequency domain. For the first 25 seconds, the participant is hunting for an object in a room. From 26 - 35 seconds, the participant walks to the door to exit the room. From 35-40 seconds, the participant opens the door. From 41-45 seconds, the participant begins walking down the hallway. From 45-55 seconds, the participant enters the stairwell and climbs stairs. From 55-60 seconds, the participant resumes walking down the hallway. . . .	89
4.7	Examples of variations within regular-gait, wrist worn FFTs. The variations shown do not cause misclassification. . . . .	93
4.8	Examples of variations across gait type and body position which resulted in misclassification of activities. . . . .	94
4.9	Stride detection accuracy as a function of window size. . . . .	95
4.10	Stride detection accuracy as a function of window size, with each stride type examined individually. . . . .	97
4.11	Irregular stride FFTs were examined to determine why the accuracy rate was low when using a 5 second window for analysis. Examples of correctly classified and misclassified FFTs are shown. . . . .	98



# Chapter 1

## Introduction

### 1.1 Overview

Healthcare is currently a reactive process which is based on symptom diagnosis and treatment. Patients primarily only seek care when sick or injured, his or her symptoms are diagnosed, and based on these symptoms, a treatment is provided based on the physician's recommendation. Technological advances in wearables have already begun improving healthcare and have the potential to reshape it entirely. Instrumenting people with wearable sensors or instrumenting their homes with devices that take daily measurements would allow conditions to be monitored in situ rather than at the clinic or hospital. Future research in wearables could reshape the healthcare system entirely by widely expanding this paradigm, allowing for more comprehensive personal health monitoring. This would allow for a reduction in the use of expensive instruments in a clinic or hospital that require an expert to operate and only provides measurements for a condition once per visit. Since clinical visits can be separated by months or years, there are significant time periods in which conditions are not monitored. In this time frame, complications can arise and problems can develop, causing extended stays and treatments for conditions which could have been handled easily with an immediate response. In-home instrumentation can improve this, reducing extended hospital stays by identifying conditions more quickly.

One example of how wearables can be used to improve healthcare can be seen in the treatment of obesity. Obesity is caused by an imbalance in energy intake and energy expenditure. Energy intake occurs when food and drinks are consumed, and energy expenditure is a combination of in-

ternal metabolism and physical activities. Using information gathered from the measurement of physical inputs (foods, drinks, etc) and physical activities (running, walking, sleeping, etc) is the first step in providing personal feedback for the treatment of obesity. A wrist worn device, capable of measuring the number of bites of food taken, or a table embedded scale, capable of determining the amount of food being consumed have been developed by our group [1, 2, 3, 4, 5]. Improvements in these technologies could provide feedback about the amount of food and drink consumed and could warn users before their energy intake becomes too high. Similarly, improvements in measuring physical activity through activity monitors, including pedometers, could help notify users when their activity levels drop. By providing objective feedback to users and encouraging increases in energy expenditure and decreases in energy intake, battling obesity can be a personalized, daily effort. If this treatment is effective, it could reduce the need for bariatric surgeries.

The healthcare industry is reaching a turning point where hospitals and clinics are beginning to include free-living health monitoring and management. Wearables are the driving force in this movement and provide physicians with a greater wealth of data regarding long term physiological signals allowing for better diagnoses. Devices already exist to monitor a wide variety of signals including blood pressure, temperature, heart rate, stress, sleep patterns and quality, cardiac fitness, and daily activity. Medication adherence tools can improve the consistency with which patients take the correct medications. These devices can provide long term measurements by sending days or weeks of data to clinics and hospitals for analysis, and they can provide immediate feedback and encouragement to users. For example, a cardiac fitness monitor could provide early warning of an impending cardiac event, sending immediate warnings to the user and hospital system. The ability of these wearable systems to gather and analyze critical physiological signals in daily living could provide proactive solutions, avoiding the development of chronic conditions requiring lengthy hospital stays for treatment. The system described requires merging of many sensors and pattern recognition algorithms in order to provide accurate feedback. These improvements would help to improve health while simultaneously decreasing healthcare costs.

The development of wearables capable of measuring physiological has already produced new sensors, devices, and even apps, which seek to take advantage of the widespread adoption of smartphones. Smartphones have a number of sensors embedded and are an easy target for developers due to the low costs of app creation and prevalence of smartphones in society. The FDA is the primary regulation body for this type of device, and it has already officially declared limitations to

oversight of smartphone apps [6]. From the FDA website, “The FDA is taking a tailored, risk-based approach that focuses on the small subset of mobile apps that meet the regulatory definition of device.” This includes apps that are intended to be used as an accessory to a regulated medical device or transform a mobile platform into a regulated medical device. The reason for this policy on regulation is likely the result of the massive number of applications being developed and used on a regular basis. Estimates in the mobile health industry indicated that over 500 million smartphone users will be using at least one healthcare application by 2015, and half of the 3.4 billion smartphone and tablet users will have downloaded mobile health applications by 2018 [7].

This means that hundreds of thousands of mobile apps may not be regulated, including those which allow users to collect blood pressure data, monitor and trend exercise activity, track quantity or quality of sleep patterns, or calculate calories burned in a workout. Many of these healthcare applications require the development of pattern recognition algorithms to provide accurate results to users. A lack of regulation in this area creates a need for developing standardized dataset, evaluation metrics, and benchmarks for evaluating the performance of new algorithms. Standard benchmarks allow algorithms’ applicability for treatment and monitoring to be evaluated and to compare performance improvements across algorithms. Effective benchmarks can also identify times and conditions under which developing technologies fail. Identifying and correcting these problem cases can provide confidence to consumers and health care professionals that the algorithms and devices developed are objective and accurate.

This work specifically considers the development and evaluation of pedometer algorithms. A pedometer is a portable device worn or carried by a person that counts each step taken. Early forms of pedometers used a mechanical switch to count steps. This was typically accomplished with either a lead ball sliding back and forth or a pendulum striking stops as it swings. Modern pedometers use micro-electromechanical systems (MEMS) called accelerometers to detect steps. Earlier pedometers used a single axis accelerometer, but the vast majority of modern devices use three axis accelerometers. This allows motion to be estimated in three dimensional space, providing more opportunity for algorithm development. Accelerometer signals are typically measured in meters per second squared or in gravities. Pedometer algorithms typically detect steps through changes in this signal, commonly looking for zero crossings or peaks. This method is particularly effective when the wearer is walking with a regular gait. In a regular gait, the wearer is walking with a consistent motion, which is characterized by a regular up and down motion in one of the accelerometer axes.

This rhythmic body motion is present during walking at multiple body locations due to regular motion in the arms, legs, and core.

Pedometer algorithms developed for commercial devices are proprietary, and there is no gold standard for developing pedometer algorithms in scientific research. Research groups have investigated how slower or faster walking can affect pedometer accuracy [8, 9, 10, 11, 12]. The location where the pedometer is worn has also been shown to affect accuracy [13, 11, 14, 15, 16]. Pedometer accuracy has also been compared between flat surfaces and stair climbing [17, 18, 19]. Prior pedometer algorithm evaluations have two major shortcomings, however. First, prior evaluations have focused only on running count accuracy (RCA), or the total number of steps taken compared to the total number of steps counted by the pedometer algorithm. Second, each pedometer algorithm is evaluated on a unique dataset collected by individual research groups, so algorithms have not been evaluated across multiple datasets. Using only RCA measures does not consider the possibility of many false positives and false negatives counterbalancing each other. A consumer’s continued use of a pedometer often depends on the believability of the device, and algorithms with many false positives and false negatives, while ultimately providing a reasonably accurate step count, may not stand up to the scrutiny of a consumer comparing small numbers of steps to the device’s provided step count. These problems indicate that a new dataset for evaluating multiple pedometer algorithms on multiple body locations, with the ability to determine specific detections (TP, FP, TN, FN), and across multiple gaits is needed.

This work is an attempt to begin building a large, free, and publicly available database of raw accelerometer signals which can be used to evaluate pedometer algorithms. The dataset collected in this work includes accelerometer data collected from multiple body locations, many participants, and recorded tens of thousands of steps. These qualities empower repeatable quantitative assessment of algorithms. This dataset would allow future researchers to test new ideas in algorithm development without having to instrument new participants and record new data. It would also facilitate algorithm performance comparisons from multiple independent researchers on a common dataset. Recommended locations for wearing pedometer devices could be informed based on the results made publicly available through use of this large standard dataset. These results could also serve as a benchmark for estimating pedometer algorithm accuracy in unregulated smartphone apps designed to provide step counts to users. We believe the described dataset could also further pedometer algorithm development by expanding evaluation to include unstructured gaits and providing

a standardized dataset that includes variations in activity, gait and body location.

## 1.2 MEMS sensors

Pedometer algorithms are developed based on the analysis of accelerometer or, occasionally, gyroscope measurements. These measurements are gathered through use of micro-electro-mechanical systems (MEMS) sensors, which are composed of very small electrical and mechanical components mounted on a single chip. MEMS sensors have low power consumption and can be operated for hours or days with continuous usage. The dataset compiled for this work was recorded using a Shimmer3 device, which contains an 3-axis accelerometer and a gyroscope. These two MEMS devices are the most common sensors used for step detection in prior works, and the collection of this raw data will allow multiple algorithms to be implemented on this dataset.

### 1.2.1 Accelerometers

Most modern pedometer devices use 3 axis MEMS accelerometers, providing acceleration measures across three axes (x, y, and z) as shown in Figure 1.1.

The accelerometer is based on the concept of the mass spring system shown in Figure 1.2. According to Hooke's law, the extension of a spring is in direct proportion with the load applied to it. Mathematically, it can be stated as Equation 1.1.

$$F = kx \tag{1.1}$$

where F is the force; k is the spring constant; x is the displacement of the springs end from its equilibrium position. In Newtons second law of motion, the force is related to acceleration shown in Equation 1.2.

$$F = ma \tag{1.2}$$

where F is force; m is the mass of the body; a is the body acceleration. Combining these two equations allows the measurement of acceleration using the displacement of a mass shown in Equation 1.3.

$$a = kx/m \tag{1.3}$$

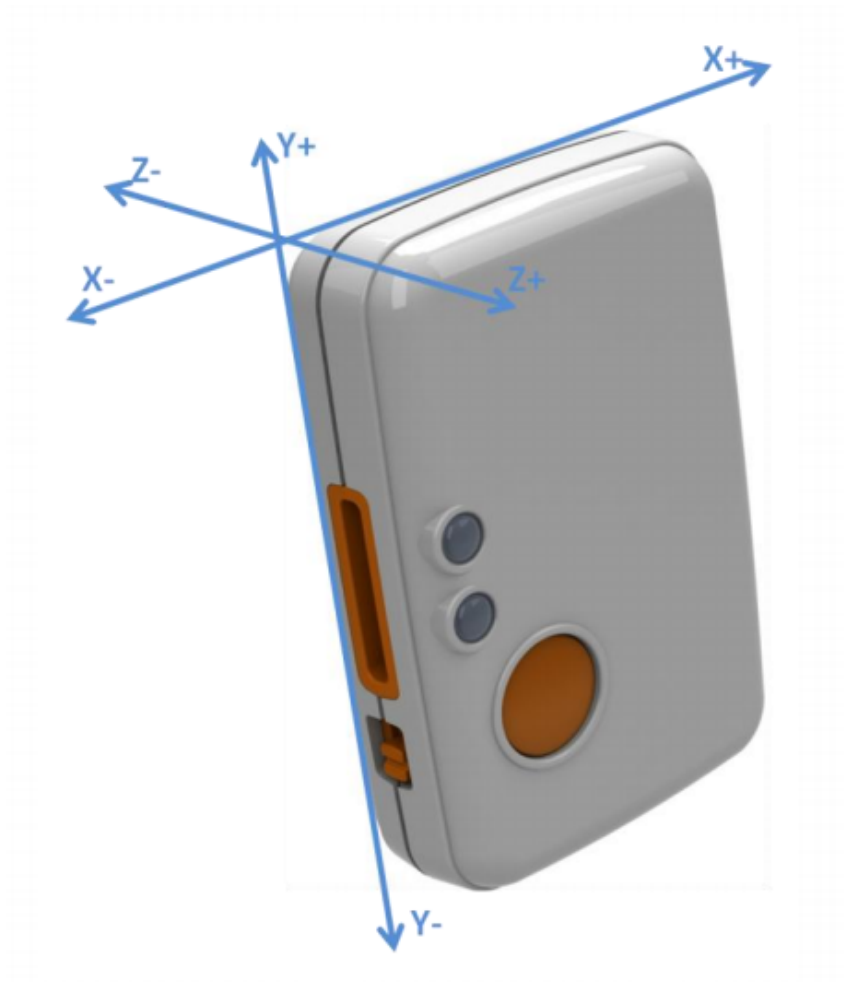


Figure 1.1: The x, y, z coordinate system used by the Shimmer3 device to track acceleration.

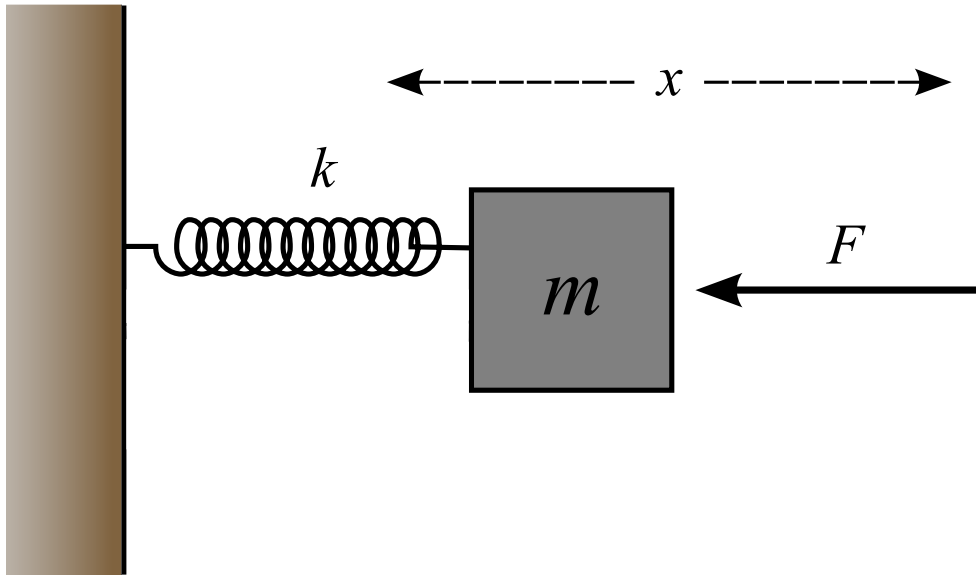


Figure 1.2: A mass spring system.

where  $a$  is the acceleration of the body,  $k$  is the spring constant,  $x$  is the displacement of the spring's end from its equilibrium positions, and  $m$  is the mass of the body.

Some pedometers use a spring-lever system based on the concepts of a spring mass system to count steps, as shown in Figure 1.3. While spring levered pedometers are still evaluated [20, 21], most pedometers use MEMS accelerometers for step detection. These devices use capacitive components to convert the physical motion from a spring mass type system into an electrical signal. The conversion process is quick, causing low enough latency to have no effect on detecting the slower physical movements in human motion. Power consumption of MEMS accelerometer devices is also very low, with a common accelerometer less than 1mW [22]. Gravity causes a constant affect on an accelerometer, appearing in whichever axis or axes are perpendicular to the earth. Accelerometers measure linear motion, excluding rotation, and when the direction of gravity is known, it can be subtracted from the accelerometer output. Gyroscopes and magnetometers are commonly used in combination with accelerometers to track rotation and calculate linear motion. A device combining both an accelerometer and gyroscope is commonly referred to as an inertial measurement unit (IMU).

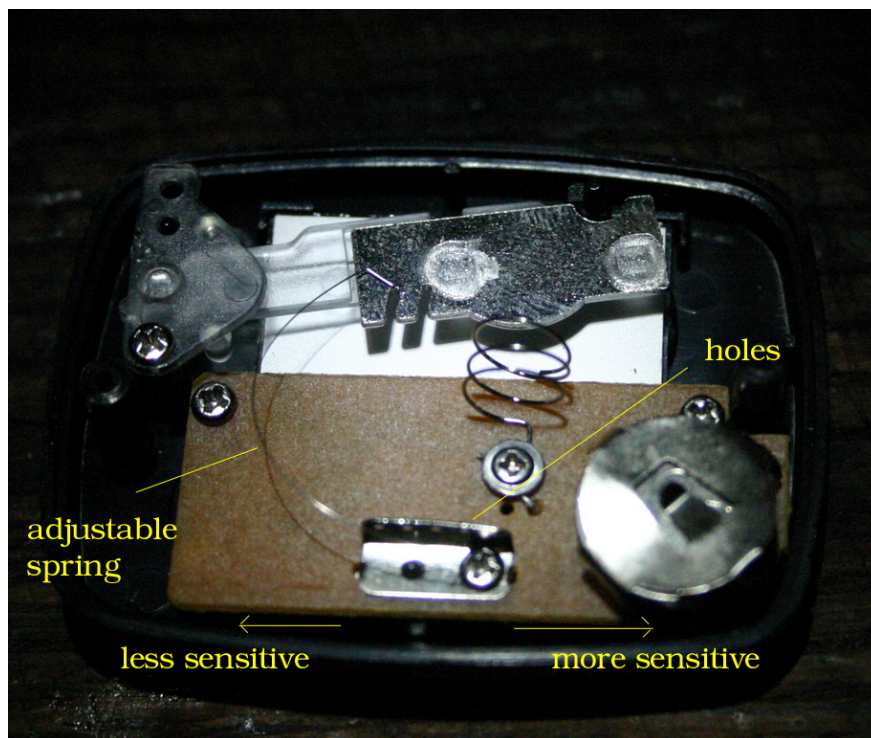


Figure 1.3: A pedometer made using the spring mass system to determine when steps occur.



## 1.2.2 Gyroscopes

With respect to pedometers, the accelerometer is the most commonly used sensor, but some algorithms have used gyroscopes to detect steps [23, 18]. A gyroscope is a device that measures rotational velocity along 3 axes. The naming of the axes as yaw, pitch, and roll were initially used to describe an aircraft rotation as shown in Figure 1.4. Yaw means left or right about an axis running up and down; pitch means up or down about an axis running from wing to wing, and roll means rotation about an axis running from nose to tail. A gyroscope measurement is based on the Coriolis effect shown in Equation 1.4.

$$F = -2m\vec{\Omega}\vec{v} \quad (1.4)$$

where  $F$  is the force,  $m$  is the mass of the body,  $\Omega$  is the angular velocity of the reference frame, and  $v$  is the body velocity. As recently as 6 years ago, a MEMS gyroscope typically consumed 10 times more power than a MEMS accelerometer [24]. This is likely a reason that many consumer activity monitors use only an accelerometer for step detection. Because devices are wearables, typically seen on wrist worn devices or in smartphones, battery life is an important consideration. Dedicated fitness monitors using only an accelerometer can routinely be used for approximately a week without needing a recharge.

## 1.3 Activity Recognition

The development of MEMS sensors, including the accelerometer and gyroscope, have provided new opportunities for activity recognition. The goal of most activity monitors is to encourage increased activity, as it has been shown that regular physical activity can aid in the prevention of multiple chronic diseases (including obesity, diabetes, cancer, and cardiovascular disease among others) across multiple age groups [25, 26, 27, 28]. This finding has encouraged the development of devices that can detect the type of activity being performed, measure energy expenditure directly, and count steps.

Accelerometers have been used to differentiate between many activities, including walking, jogging, running, climbing up or down stairs, sleeping, eating, sitting, and standing, among others [29, 30, 31, 32, 33]. Most of this research takes place in a controlled setting, specifically focusing on

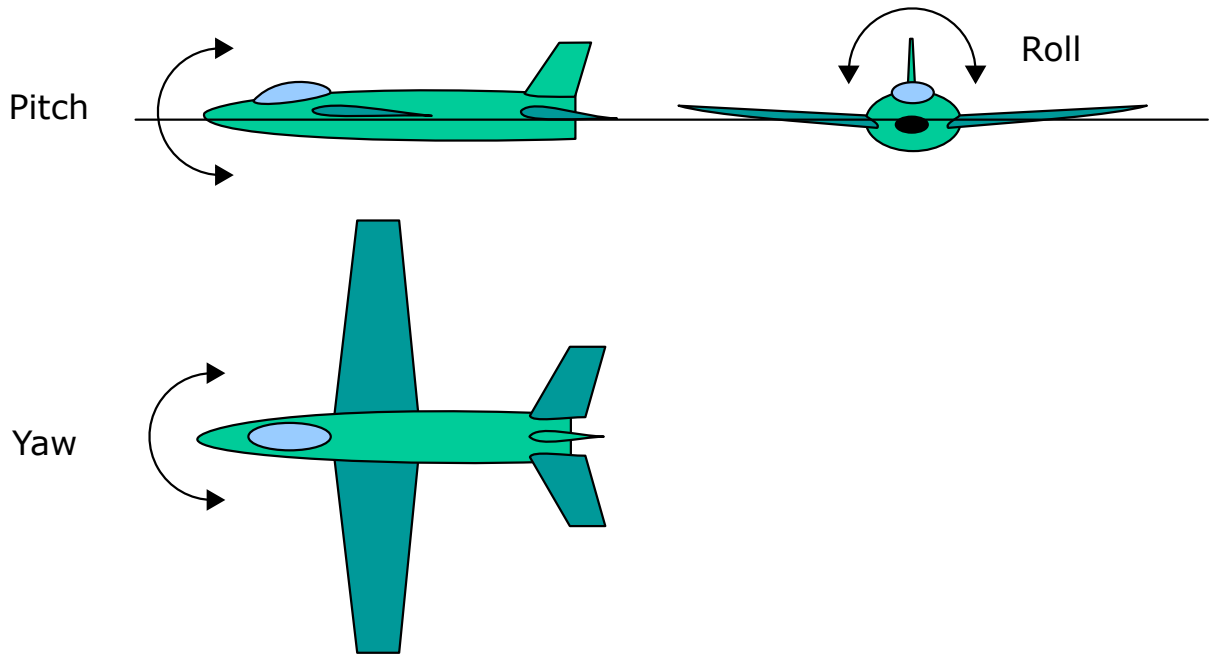


Figure 1.4: The roll, pitch, and yaw system used to define the measurement from a gyroscope.

differentiating between activities with regular gaits (walking, jogging, running, climbing stairs, etc) and activities with little body motion (sleeping, eating, sitting, standing, etc). When regular gaits are examined, the accelerometer signal is typically very periodic, with steps being visible in the cyclic nature of the signal, as seen in Figure 1.5. There are some rare studies which examine behaviors with more irregular motion patterns including mowing, raking, shoveling, vacuuming, sweeping, and stacking groceries [34]. The majority of research papers exploring various types of activity recognition use accelerometers while only a few utilize gyroscopes. Both sensors have advantages and disadvantages, but it is assumed that the increased power consumption historically required by gyroscopes led to a faster spread of accelerometer sensors and algorithms [35]. Recent advances in technology have since significantly reduced power consumption requirements of gyroscopes. An individual accelerometer, the ADXL345 (<https://www.sparkfun.com/datasheets/Sensors/Accelerometer/ADXL345.pdf>) draws  $145 \mu\text{A}$  of current for high sample rates ( $>100 \text{ Hz}$ ) or  $40 \mu\text{A}$  of current at low sample rate ( $<10 \text{ Hz}$ ). The current draw in a modern IMU containing both an accelerometer and gyroscope, the KXG07, is  $200 \mu\text{A}$  in normal mode, or  $600 \mu\text{A}$  in high resolution mode. (<http://www.kionix.com/product/KXG07>) From this comparison, it is clear that the power consumption of gyroscopes is no longer a factor of 10 greater than accelerometers, but it is probably closer to a

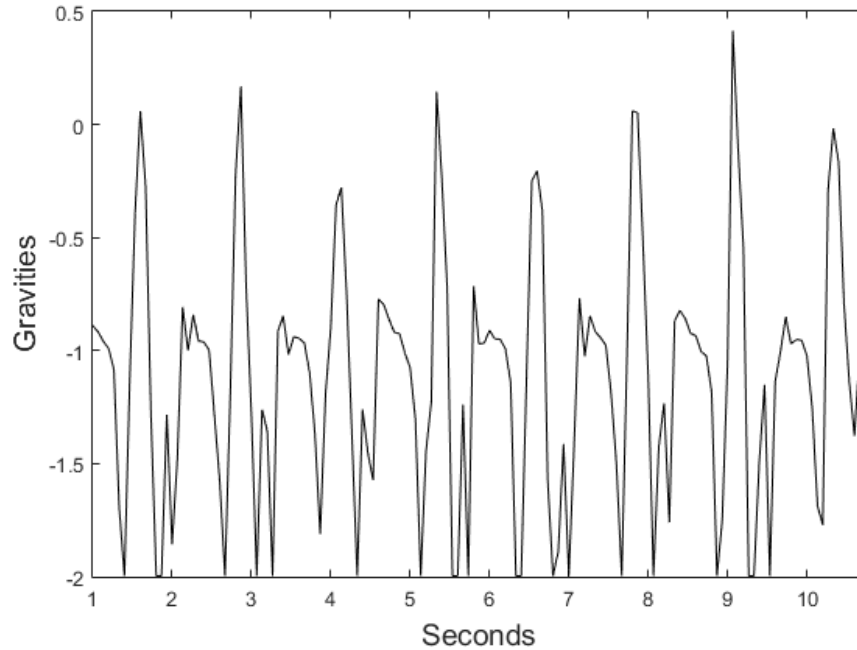


Figure 1.5: Sample accelerometer signal typical in regular gait analysis.

factor of 3.

With respect to obesity, the critical factor in activity recognition is the determination that energy expenditure (EE) is greater than energy intake (EI). In order to estimate this deficit, it is critical to develop tools capable of measuring EE for multiple activities. The most accurate method of determining EE is by using a calorimetry chamber, in which subjects live in a fully controlled environment. In the calorimetry chamber, a subject's EI can also be measured by using a technique called bomb calorimetry for all foods provided to the subject. Calorimetry chambers are very expensive and also extremely impractical for widespread use over time [36]. The most accurate technique currently available in a more realistic setting for measuring EE is the use of doubly labeled water [37]. Over a set period of time, frequently one week [38], the average metabolic rate of a subject is measured by using water in which the hydrogen and oxygen are replaced with deuterium and oxygen-18. By replacing these elements in the water consumed, the decay of the isotopes consumed is measured through use of daily saliva, blood, and urine samples. This technique, while applicable in regular daily life, is still very expensive and requires technical expertise to take samples [39]. Indirect calorimetry is another method for measuring EE which measures inspired and expired gas flows, volumes and concentrations of oxygen and carbon dioxide, and through comparison of these

values, estimates EE. The use of indirect calorimetry for EE estimates is becoming increasingly popular in laboratory settings because of its non-invasiveness and ease of use [40, 41, 42]. Even this method, however, requires users to wear a mask, which limits its ability to be adopted and used at a consumer level.

Step counting devices are quickly becoming very common at a consumer level, with multiple smart phone applications and wearable devices including smart watches and pedometers made by companies including Fitbit, Samsung, Garmin, etc reaching an unprecedented level of consumers. Evidence supporting the value of pedometer usage indicates that “the use of a pedometer is associated with significant increases in physical activity and significant decreases in body mass index and blood pressure.” [25]. Many of the commercial pedometer devices developed also provide distance estimates, energy expenditure estimates, heart rate estimates, and a number of other health measurements. While use of pedometer devices has been correlated with increased step counts [43, 44], features including energy expenditure estimates have shown mixed results [8, 45]. This is likely the result of multiple factors, including that pedometers do not quantify stride length or total body displacement [46, 47], however, with improved algorithms, stride length could be more closely estimated and closer estimates of energy expenditure could be achieved.

## 1.4 Pedometer Algorithm Evaluation

Table 1.1 summarizes previous works on pedometer evaluation. We consider several criteria in this summary, including the number of subjects, number of steps, gait types, sensor locations, method for determining ground truth step counts, and evaluation metric. This section explores the importance of each of these criteria and the shortcomings of previous works.

Prior algorithms evaluating pedometer performance have a wide variety in the number of participants being evaluated. Many evaluate only one participant, and a stark contrast can be seen when comparing the single participant evaluated in [68] to the 140 participants evaluated in [69]. Testing on a large number of participants is important because everybody walks differently. Height, age, weight, health, posture, and other individual factors may affect how a pedometer algorithm should be tuned to provide optimal performance. Table 1.1 shows that most studies test their algorithm on fewer than ten subjects and many studies just test one subject. This can demonstrate the potential of an idea but falls short of providing a comprehensive, generalizable test that can be

Cite	# Sub	# Steps	Gait types	Locations	GT method	Metric
[48]	9	2,700	Walk	Pocket, hand	Self count	RCA
[49]	No Info	500 per sub	Walk(3), run	Wrist(3), hand	No info	RCA
[16]	No info	11,700-35,100	Walk and run	Pocket(3), hand, bag	No info	RCA
[50]	6	2,400	Walk,run,stairs	Hip	No info	RCA
[51]	5	4,750	No info	Chest	OMRON	RCA
[52]	10	4,516	Walk	Pocket(2), bag,hand, wrist	Self count	RCA
[53]	30	189,210	Walk, run	Hand(3), pocket, bag(2), arm	No info	RCA
[54]	No info	No info	No info	No info	No info	No info
[55]	No info	300 per sub	Walk	Hand(6), pocket(3), bag(3)	No info	RCA
[56]	20	155,000	No info	Hand	No info	RCA
[14]	No info	200 per sub	Walk, run	Hip,pocket, bag	Self count	RCA
[18]	10	5,468	Walk(2), incline, stairs	Leg	No info	RCA
[57]	4	No info	Walk	Hand(3),bag, foot	Foot sensor	Individual
[58]	27	No info	Walk multiple	Pocket, bag(2)	Video	RCA
[59]	1	2,480	Walking(5)	Pocket	No info	RCA
[60]	3	No info	Walk outside	Pocket, hand	No info	No info
[61]	6	No info	Stationary, walk	Pocket(3), bag, hand(2)	No info	Individual
[62]	3	1,800	Walk(2)	Hip	Observer	RCA
[63]	3	No info	Walk	Chest, leg	No info	No info
[64]	No info	No info	Free living	Hand(3), pocket, bag	No info	No info
[65]	No info	No info	No info	Hip	No info	No info
[19]	3	1,384	Walk, run, stairs	Right foot	No info	RCA
[66]	6	No info	Walk	Ankle	No info	No info
[67]	No info	No info	No info	No info	No info	No info
[68]	1	17,367	Urban, stairs	Pocket	Nike+	RCA
[69]	140	20m per sub	Walk	Hip	Video	RCA
[70]	1	No info	Walk	Foot	No info	No info
[71]	5	1,000	Walk,sloped, stairs	Ankle	No info	RCA
[72]	No info	No info	No info	Head	No info	No info
[73]	49	3,650	Walk	Hip	Observer	RCA
[74]	1	300	Walk	Ankle	No info	RCA

Table 1.1: Summary of evaluations performed for pedometer algorithm development. (Numbers in parentheses indicate variations within the location or pace indicated. For example, Walk(3) indicated that 3 different walking paces were examined and Hand(3) indicates that the sensor was held in the hand in three different orientations or postures)

repeated. Lee et. al. evaluate 30 participants in [53], however, the method of determining ground truth steps is not described, the surface worked on is not described, and the positions examined only include smartphone positions (various methods of being held in a hand or place in a bag or pocket are examined), failing to test locations where fitness tracking devices are commonly worn (on the wrist, attached to the hip, or around the ankle). Lin et. al. evaluate 20 participants in [56] but do not describe the pace or gait used, only evaluate a hand held device, and do not describe the method used for determining ground truth step count. Brajdic et. al. evaluate 27 participants in [58] but do not specify the number of steps taken by each participant and only examine devices in the pocket or in bags. Marschollek et. al. evaluate 140 participants in [69], but only examine 20 meters of walking per participant, only examine walking in a hallway, and only have sensors on the hip of participants. Similarly, Ichinoseki-Sekine et. al. evaluate 49 participants, but only record each participant walking approximately 20 meters [73]. This work describes a dataset collected from 30 participants and future work may include expanding the dataset to hundreds of people.

Similar to subject count, the number of steps tested on is an important statistic for pedometer evaluation. While a large number of participants is important to ensure that a pedometer algorithm is not specifically tuned for a single person’s gait, a large number of steps per participant is important for ensuring an algorithm can be accurate for a variety of step types within a participant. Even for one person, a step taken walking down a hallway can look different than a step taken walking on a nature trail. The first step in walking down a hallway looks different than the twentieth step taken down the same hallway which looks different than a step taken when turning. A large number of steps is needed to capture each of these unique step types within an individual’s gait. The number of steps per participant analyzed in prior algorithms can range from 200 to 17,367 per participant. Some evaluations do not specify the number of steps taken because the focus of the paper is on detecting the distance traveled. These algorithms use step detection as a component of their distance algorithms (typically combining number of steps with estimated stride length), but because step detection is such a critical component of their algorithms, it is still important to report the number of steps evaluated. Lin et. al report 20 participants walking 50, 100, 200, 400 and 800 steps five times each, or 7,750 steps per participant, totaling to 155,000 steps [56]. However, in this work, the method used to determine a ground truth step count for each of these tests is not reported, so it is difficult to verify the accuracy of the results. Each participant should take enough steps to provide an accurate characteristic for evaluating their stepping motion in each type of gait

analyzed. There is no defined requirement for this metric, but it should be enough to capture all of the step motions included in each gait studied. The dataset collected for this work includes over 60,853 steps and may be significantly expanded in future work.

Another critical component in evaluating pedometer algorithms is determining how changes in gait affect the accelerometer signal and the accuracy of the algorithm. When thinking about the problem of step detection, the first thing that comes to mind is intentional exercise, including running or walking on a treadmill. While this type of gait is common when considering increasing daily step count, a significant number of steps occur from normal daily activities. While these steps occur in a regular, repeating pattern, the steps taken when walking around a house or an office, performing chores, or walking around a room are interrupted and more challenging to identify from accelerometer signals. Many studies evaluating pedometer accuracy have focused on how changes in the rate of motion, inclines, and stairs affect pedometer accuracy when gait is regular. None of these studies have focused on gaits which are interrupted by pivots, shuffle steps, or brief pauses. In other words, the steps taken in daily activities are not evaluated in any of the algorithms described in table 1.1. In fact, many studies have gone out of their way to avoid these transitional steps. Brajdic et al. evaluate several pedometer algorithms across multiple rates of motion, but specifically “excluded all traces where there was not a clear step count,” where unclear step counts were due to shuffling or unclear steps [58]. These unclear steps are included in some studies, like the one performed by Lu et. al. which combine an accelerometer, gyroscope, and pressure pad to identify when participants were sitting, walking, running, or biking [75]. This study was focused on activity recognition and step detection, but only reported step detection accuracy for the walking condition. Recognition of all gaits, including the more challenging irregular gaits present in daily activities is essential in evaluating how effective a pedometer algorithm will be when applied in a natural environment. This work describes a dataset which is focused on capturing regular, semi-regular, and irregular gaits which are more representative of the variety of walking present in daily living.

Pedometer accuracy has been evaluated when pedometers are worn on multiple body locations. Evaluating pedometers on multiple body locations is particularly important when considering the wide variety of activity tracking devices available to consumers. Pedometer devices can be worn on the wrist, clipped to shoes or clothing, held in the hand, or kept in a bag, just to name a small subset. Consumers may not comply with the recommended locations for pedometer devices, increasing the number of places these devices can be worn. It is important to evaluate pedometers

at various locations in order to determine which location provides the highest accuracy for each algorithm as well as to evaluate how much of a decrease in accuracy is realized when the position on which the pedometer is worn changes. Throughout the list of prior algorithms in Table 1.1, a number of locations have been evaluated, including the head, chest, arm, hand, wrist, pocket, hip, leg, ankle, foot, and in a bag. There has been a large focus on locations where smartphones may be located, including multiple orientations that could occur while the phone is in use (texting, calling, etc) [53, 55, 57, 61, 64], so many variations of hand, pocket, and bag orientations have been studied. In addition, within each location, it is possible for an activity monitoring device to be rotated or positioned in a number of ways. This problem of evaluating how position affects pedometer accuracy is challenging based on the sheer number of possibilities it presents. It is possible that the problem could be simplified by taking the magnitude of all three axes of accelerometer signals, reducing the effect of orientation of the device. In addition, it is likely that that multiple locations can be represented by one characteristic accelerometer pattern. For example, the hip and pocket will have similar motion patterns, though a greater level of noise may be present for a device in a pocket. This work considers the wrist, hip, and ankle as locations which represent three significantly different accelerometer patterns and may expand the variety of locations recorded in future work.

Ground truth measurements provide an external benchmark of step count against which algorithm measurements of step count can be compared. For regular gaits, counting steps can be done by having a participant count their own steps or by having another observer count the steps. This ground truth metric is prone to human error, particularly as the number of steps taken increases. In addition, there is no method for reviewing an experiment to verify if the reported step count is accurate. These difficulties are exacerbated when irregular gaits are considered because in irregular gaits it can become challenging to determine if a motion should be considered a step, even with video review. Of the papers examined, fewer than one third specified the method used to determine a ground truth step count. Of those that did specify a ground truth method, three used a step count provided by a reference sensor as a ground truth for evaluation of their proposed algorithm. Chien et. al. used a commercial Omron pedometer to validate the step count determined by their algorithm [51]. Mladenov et. al. use a Nike+ sports kit, a commercial device which uses a shoe-mounted accelerometer to detect steps, as a reference step count for validation [68]. Susi et. al. record data with two IMUs, with one IMU on the ankle used as a reference for the accuracy of the second IMU [57]. This type of bootstrapping assumes the reference sensor is accurate, which is a



bold assumption to make. Three additional studies specify that the subject counts their own steps. This is likely a common method for the types of studies performed because of the ease of counting steps when continuously walking. Two studies indicate that the experimenter counts the steps of the participant. This method is also likely very common in the experiments which do not specify their method for determining ground truths, and it suffers the same shortcomings as the self count method. For natural gaits, including complex steps, accurate review and ground truthing should be done with video review. Only two prior studies indicate that video review was used as the metric for gathering a ground truth step count. One of these two studies doesn't provide the number of ground truth steps recorded and only examines sensors worn in a pocket or in a bag [58]. The other study to use video review only examines walking in a hallway with a sensor worn on the hip [69]. This work uses video review to determine the ground truth steps taken, and due to the challenges in irregular gait, it would have been impossible to determine an accurate step count without use of video.

Pedometer accuracy can be evaluated according to a number of different metrics. The most common evaluation metric used is running count accuracy (RCA), or the number of steps reported by the device compared to the number of steps taken by the wearer. The equation used for this form of evaluation is shown in Equation 1.5

$$RCA = \frac{(d - a)}{a}(100\%) \quad (1.5)$$

Where  $d$  is the number of steps detected by the algorithm and  $a$  is the actual number of steps (ground truth). This equation is a simple accuracy equation which divides the difference between detected steps and ground truth steps by the number of ground truth steps. Using RCA as the evaluation metric leaves the possibility that an algorithm would be evaluated as very accurate if the number of false positives and false negatives is roughly even. In practice, this may allow for a roughly accurate step count over long periods of time, but inaccurate results for small numbers of steps or gait changes. The alternative to RCA is to evaluate pedometer algorithms based on individual step detections. This type of evaluation allows the number of true positives, false negatives, and false positives to be determined. True negatives should likely be excluded from calculations because the number of steps, compared to the number of accelerometer readings, which typically occur between 15-100 Hz, is very small. Since the vast majority of data points will be negatives, identifying a true negative is a poor

metric for determining the effectiveness of a pedometer algorithm. Of all the algorithms in table 1.1, only two evaluated pedometer accuracy based on individual step detections. Susi et. al. only evaluate 4 participants, do not specify the number of steps recorded, and only examine the ankle [57]. Rai et. al. only record six participants, do not specify the ground truth method used, and do not provide the total number of steps taken [61]. The use of the individual step detection evaluation metrics could be particularly impactful when considering consumer believability. Believability, with regard to pedometers, is the users belief that the step count reported by the pedometer matches the actual number of steps they have taken. Consumers may perform limited and poorly controlled tests of pedometers to determine which brand to purchase, which body location to wear it on, and their overall confidence in the reported measure. The nature of these limited and poorly controlled tests requires that the pedometer algorithms employed in commercial products are robust enough to be accurate over a small number of steps, not just accurate over the course of thousands of steps. Greater believability encourages continued use, while lower believability may cause the device to fall out of use. Because video review is used in this work, step times could be identified at specific time indices in the accelerometer data, facilitating individual step detection and evaluation.

In summary, the number of participants, number of steps, gaits analyzed, positions monitored, method used to determine ground truth, and evaluation metrics are all important factors to consider when evaluating pedometer algorithms. None of the papers summarized in Table 1.1 provide ideal evaluation methodologies. In addition, each algorithm is tested on a unique dataset, preventing direct comparison of each algorithm. While some evaluations have been performed on step detection while climbing stairs or walking on a slope, none of the papers evaluate step detection on interrupted gaits commonly seen in natural walking. In addition, the vast majority (28/30) of the papers use running count accuracy (RCA) to evaluate pedometer accuracy rather than a more detailed individual step detection evaluation. Because of the inconsistencies in evaluation methodologies, it is clear that a gold standard dataset for pedometer evaluation is necessary in order to provide a reliable assessment of pedometer accuracy.

## 1.5 Contribution and Novelty

To our knowledge, this is the first attempt to create a large raw accelerometer dataset for pedometer algorithm evaluation. This work describes a dataset collected from 30 participants and

future work may include expanding the dataset to hundreds of people. The dataset also includes 60,853 steps, averaging just over 2,000 steps per participant. This work is the first to our knowledge to count steps in interrupted gaits representative of those performed in daily activities. We recorded data at three positions, wrist, hip, and ankle, which provide three distinct accelerometer signals for locations at which activity monitors are commonly worn and could be expanded in future work. In order to provide accurate step counts, particularly in activities with interrupted gaits, manual review of video recordings was performed. By synchronizing the video with the accelerometer signal, individual steps are detected and marked in the data, allowing for algorithm evaluation at an individual step level. Combining all of these features creates a comprehensive dataset capable of evaluating multiple pedometer algorithms on natural gaits.

Prior studies evaluating the performance of pedometer algorithms, as shown in Table 1.1, have wide variation in participants, steps, types of gait, sensor locations, ground truth methods, and evaluation metrics because there is a lack of a gold standard. In addition, prior studies have not evaluated pedometer accuracy across motions including starts, stops, pivots, and shuffle steps common in natural walking. Whereas at least one study in activity recognition [34] has examined activities with irregular and interrupted gaits, the same task has not been attempted in the development of step counting algorithms. It is possible that these steps and motions have been avoided due to the challenges in detection, but they are critical, particularly when considering the believability of the device. A dataset including multiple gaits, multiple sensor locations, activities which encourage free living motions, and video recording for accurate ground truth would allow many pedometer algorithms to be analyzed and compared on the same dataset. This paper presents a pilot dataset which can be used as a starting point for this type of database for pedometer evaluation.

## Chapter 2

# Data Collection and Evaluation of Steps, Shifts, and Fitbit

### 2.1 Methods

This chapter describes the process used to develop and analyze a dataset for pedometer algorithm evaluation. The dataset provides raw data collected using three Shimmer3 devices from the wrist, hip, and ankle during three activities designed to elicit regular, semi-regular, and unstructured gaits for 30 participants. Each participant's lower extremities were recorded through each experiment, and the video was synchronized with the accelerometer data through use of a software tool developed for the task. Through manual review of the synchronized video, 60,853 steps were manually annotated. Through this process, it was determined that, in addition to traditional steps, a second class of steps must be defined based on differences in behavior, accelerometer signal, and, likely, in energy expenditure.

Preliminary analysis and results found in this process are described in a paper published in the IEEE International Conference on Bioinformatics and Biomedicine [76]. In this work, the methods for creating the dataset are described, and the identification of the two types of motion, called steps and shifts, are discussed. The frequency with which steps and shifts appear in each of the three gait types is analyzed and possibilities for future work are discussed. The commercial Fitbit device worn by each participant is evaluated against step tallies collected from each gait type

examined in the dataset.

### 2.1.1 Data Collection

Participants were recruited via email and provided a \$20 Amazon gift card for their participation. In all, thirty participants were selected, including 15 females and 15 males. Each subject provided informed consent and filled out a Physical Activity and Readiness Questionnaire (PAR-Q) [77]. The PAR-Q is a form used to assess whether participants should check with their doctor before performing physical activity. A copy of the PAR-Q used in this data collection can be found in Appendix C. Participants provided height, weight, and gender information. The study was approved by the Clemson University Institutional Review Board for the protection of human subjects (IRB Number: IRB2017-048).

The primary goal in designing the experiment was to elicit regular, semi-regular, and unstructured gaits for each participant. In order to accomplish this, we designed three activities, each designed to elicit one of the three gait types. Each activity was designed to take approximately 10 minutes. The activities took place on the campus of Clemson University. The regular gait activity was performed on a relatively flat segment of sidewalk behind Cooper Library. The path is roughly 350 meters in length, and participants were instructed to walk two laps around the path (for a total of approximately 700 meters) at their typical walking pace.

For the semi-regular gait activity, participants were asked to perform a scavenger hunt. Four toy footballs were placed in four different rooms throughout Riggs Hall and placed to require participants to look around and behind objects, though participants did not have to open any drawers or cabinets. Riggs Hall is five stories tall, four of which were used throughout the experiment. Participants were instructed which room to walk to and that the footballs were not located inside any objects, just between or behind things. This scavenger hunt required participants to walk along hallways and up and down stairs to travel between rooms. In addition, the activity encouraged starting and stopping motions, pivots, and shuffle steps as objects were hunted for.

The unstructured gait activity was designed to simulate the motions used when making a meal. The simulation was based on constructing a Lego toy. The Lego was constructed at a central location in a room in Riggs hall, and pieces for the Lego were distributed in 12 numbered bins in locations around the room. Participants were asked to construct as much of the Lego as possible with the pieces from each bin, then collect another bin, return to the central table, and continue

constructing the Lego until complete. Through this process, the Lego construction simulated mixing ingredients when cooking, and retrieving bins simulated gathering additional ingredients. This simulation generally exemplifies any household chore in which there is significant wrist motion with relatively few interspersed steps in which no regular gait is achieved. These three activities represent gaits common in natural daily living and encourage steps taken outside of the regular gait that is commonly studied in previous works.

Throughout the full experiment, participants were outfitted with three Shimmer3 devices for recording motion data, shown in Figure 2.1. The device can be turned off or on through use of a switch and the LEDs indicate the state of the device. The Shimmer3 is a wearable sensor weighing 23.6 grams with two accelerometers, a gyroscope, a magnetometer, and an altimeter. The device has connectors which allow for expansion to include ECG, EMG, GSR+, and other sensors, but no other sensors were attached for the collection of this dataset. An SD card is used to store data, and data can be transferred to a CP for analysis through a base station, shown in Figure 2.2. The accelerometers, gyroscope, magnetometer, and altimeter all have a selectable range for recording data. The two accelerometers allow for the selection of a low noise but smaller magnitude recording or a higher noise but wider range. The accelerometers report data in gravities, and the wide range accelerometer can record at  $\pm 2g$ ,  $\pm 4g$ ,  $\pm 8g$ , and  $\pm 16g$ , with a sensitivity of 1000 LSB/g at  $\pm 2g$ . The low noise accelerometer records at  $\pm 2g$  with a sensitivity of  $600 \pm 18$  mV/g. The low noise accelerometer was used in collecting this dataset. The device was set up to record with a sample rate of 15 Hz.

Throughout the data collection process, each participant wore 3 Shimmer3 devices, located on the wrist, hip, and ankle. These locations were selected based on two key factors: 1) the common locations on which activity monitors are worn and 2) examining locations which provide different motion characteristics. Most commonly, activity monitors, including those made by Fitbit, Garmin, and Omron, among others, are designed to be worn as a watch due to consumer popularity and acceptance. With the increasing popularity of smart watches, including the Apple Watch, the wrist is a critical location for examining pedometer algorithm accuracy. Another common commercial pedometer design is a clip-able device which can be attached to a belt, shoe, or a variety of other locations. Based on this design, we had participants wear a Shimmer3 device around their waist, like a belt, and around their ankle, as if attached to a shoe or sock. This analysis led us to select the non-dominant wrist, hip, and non-dominant ankle for sensor placement. These locations also

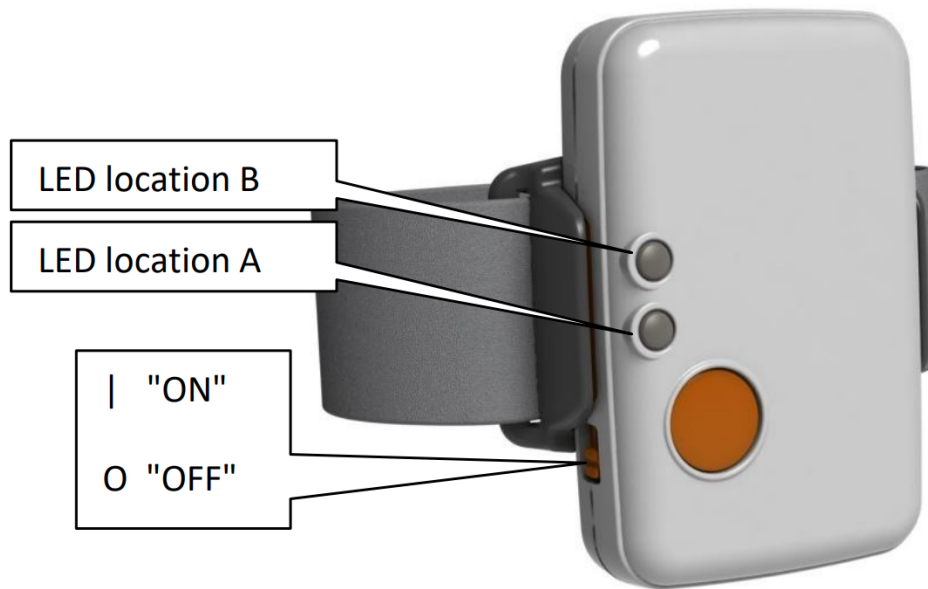


Figure 2.1: The Shimmer3 device used for motion data collection.

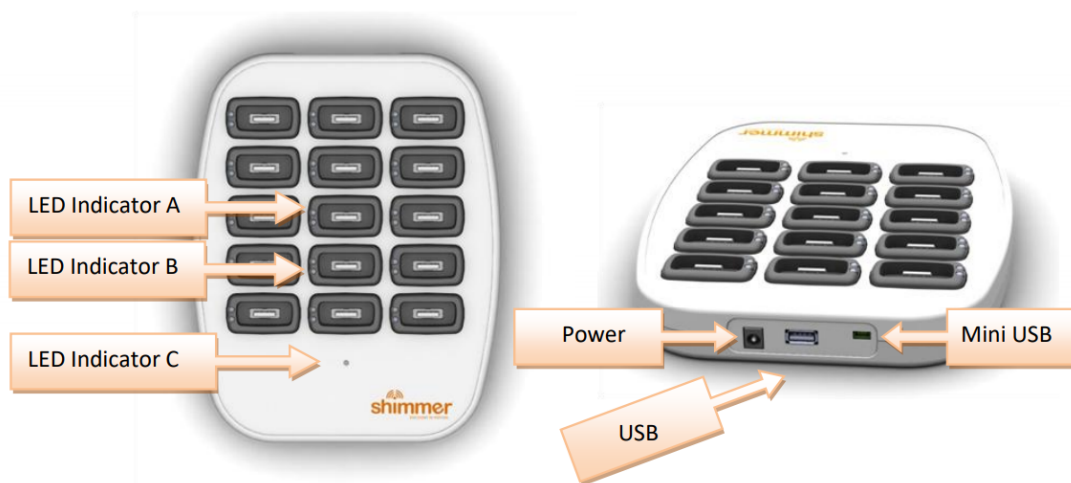


Figure 2.2: The base used to connect the Shimmer3 device to the PC and transfer data.

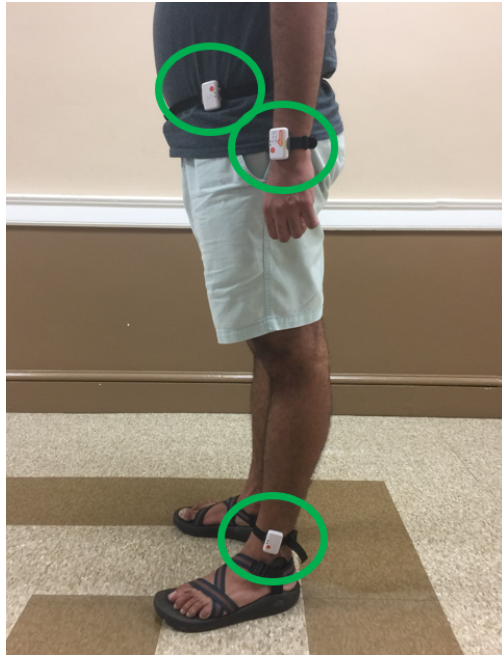


Figure 2.3: A participant wearing the three Shimmer3 devices, one each on the wrist, hip, and ankle.

produce a different accelerometer signal, as each moves in a different way while walking. The hip moves up and down with each step as weight is shifted from one foot to the other. The wrist moves with a pendulum type motion while also moving up and down with the core of the body. The ankle has periods of high activity as the instrumented leg is moved and periods of rest as the other leg takes a step. Each participant was requested to position the sensor naturally within each location. An example of a participant with all three Shimmer3 sensors attached is shown in Figure 2.3.

In order to allow for accurate ground truth step identifications, participants' lower extremities were recorded through each activity. In order to facilitate this, an experimenter followed behind the participant and recorded their legs and feet as shown in Figure 2.4. The purpose of the video is to show when a participant's feet are lifted and touch the ground (i.e. taking a step). A portion of the legs and lower torso are also visible, but participant faces are not recorded to help preserve privacy. This process produced video similar to the still image displayed in Figure 2.5. This process is critical for identifying every step taken, particularly in the semi-regular and unstructured gait activities. The iPhone4 was used for video recording. The video was collected at 30 frames per second at a resolution of 1280x720.

In addition to the Shimmer3 device, participants each wore a Fitbit Charge 2, as shown in





Figure 2.4: The procedure used in order to record all steps taken by the participant through each activity.

Figure 2.13 on their non-dominant wrist, directly adjacent to the Shimmer3 device at that location. At the start of each experiment, the participant's height, weight, handedness, age, and gender were recorded. The height and weight were updated in the Fitbit settings and the Fitbit device was synchronized before the experiment began. Each activity was described to the participant, and an example of the toy football that would be searched during the semi-regular gait activity was provided. The participants then were asked to put on the Shimmer3 devices and the Fitbit Charge 2. The only requirement when attaching the Shimmer3 devices was that the red button used to turn on and off the device, as seen in Figure 2.1, was oriented towards the hand for the wrist worn device and towards the ground for the hip and ankle worn devices. Beyond this requirement, participants were free to position the sensors as desired on each body location. Each participant was walked to a starting location before each activity, a corner of the path for regular gait, an entrance to Riggs Hall for semi-regular gait, and the table at which the Lego is assembled for unstructured gait. Once in position, the Shimmer3 devices were turned on and the number of steps displayed on the Fitbit device were recorded. Once all Shimmer3 devices were confirmed to be on and recording, the participant began the activity.



Figure 2.5: The image produced by the recording process.

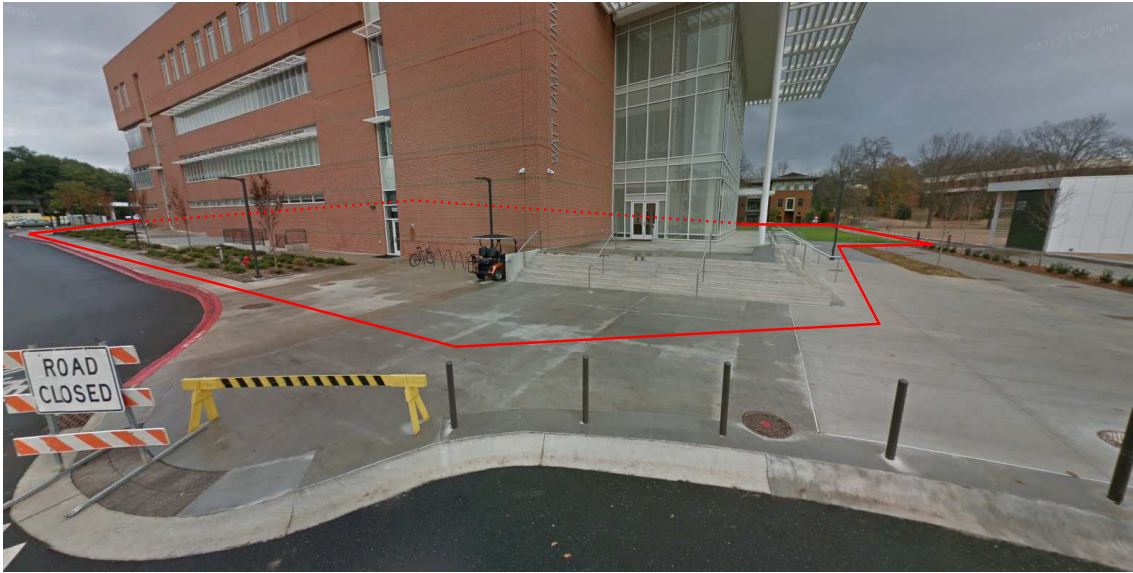


Figure 2.6: The path followed during the regular gait activity is identified with a red line (the portion of the sidewalk hidden by the building is indicated by a dashed line).

Before and during each activity, instructions were provided to each participant. Before beginning the regular gait activity, the sidewalk path to be followed was described and pointed out to each participant. The walking path is a sidewalk located behind Cooper Library, as outlined in Figure 2.6. During the first lap around the path, participants were reminded about each upcoming turn with enough time to ensure the turn could be made without a pause. Participants were instructed to continue walking through the end of the first lap, continuing on to the second lap with no pauses.

Before beginning the semi-regular gait activity, participants were reminded that their task was to locate four toy footballs, identical to the one in Figure 2.7, shown to them before beginning the activity. They were instructed that the footballs would not require anything in the rooms to be moved or opened. Additionally, they were told that once a room was located, they were to wait in front of the door as the experimenter unlocked the door for them. Between rooms, participants engaged in some regular gait walking as they traveled down hallways similar to the one shown in Figure 2.8. Once the door was open and the experimenter was in position to continue recording, participants were allowed to enter the room and begin searching for the toy football. An example of one of the rooms searched by participants is shown in Figure 2.9. Once a football was found, participants were given the next room number and were allowed to continue to the next item. Upon



Figure 2.7: One of the toy footballs that participants searched for during the semi-regular gait activity.

finding the final (fourth) item, participants stopped in front of room 305 in Riggs Hall, where the activity ended.

For the unstructured gait activity, participants were brought to Riggs 309 and shown the instruction booklet for the Lego toy to be built. The original instruction booklet, nearly assembled Lego, and a bin containing the last set of pieces are shown in Figure 2.10. Participants were allowed to see an image of the completed Lego. The procedure for constructing the Lego was then described as follows. Participants were to collect one bin of Lego pieces, return to the construction table, assemble the pieces, and collect the next bin of pieces, as shown in Figure 2.11. Participants were provided with a printed copy of the instruction booklet, with a handwritten number in the top right of each frame, as demonstrated in Figure 2.12, corresponding to the bin number containing the pieces necessary to complete the instruction. Participants were instructed that they were to collect only one bin of pieces at a time and must complete the assembly of those pieces before retrieving another bin. In total, the 89 pieces necessary to complete the Lego were divided across 12 numbered bins. The activity was complete when the Lego toy was fully assembled.





Figure 2.8: One of the hallways in Riggs Hall that participants traveled during the semi-regular gait activity.

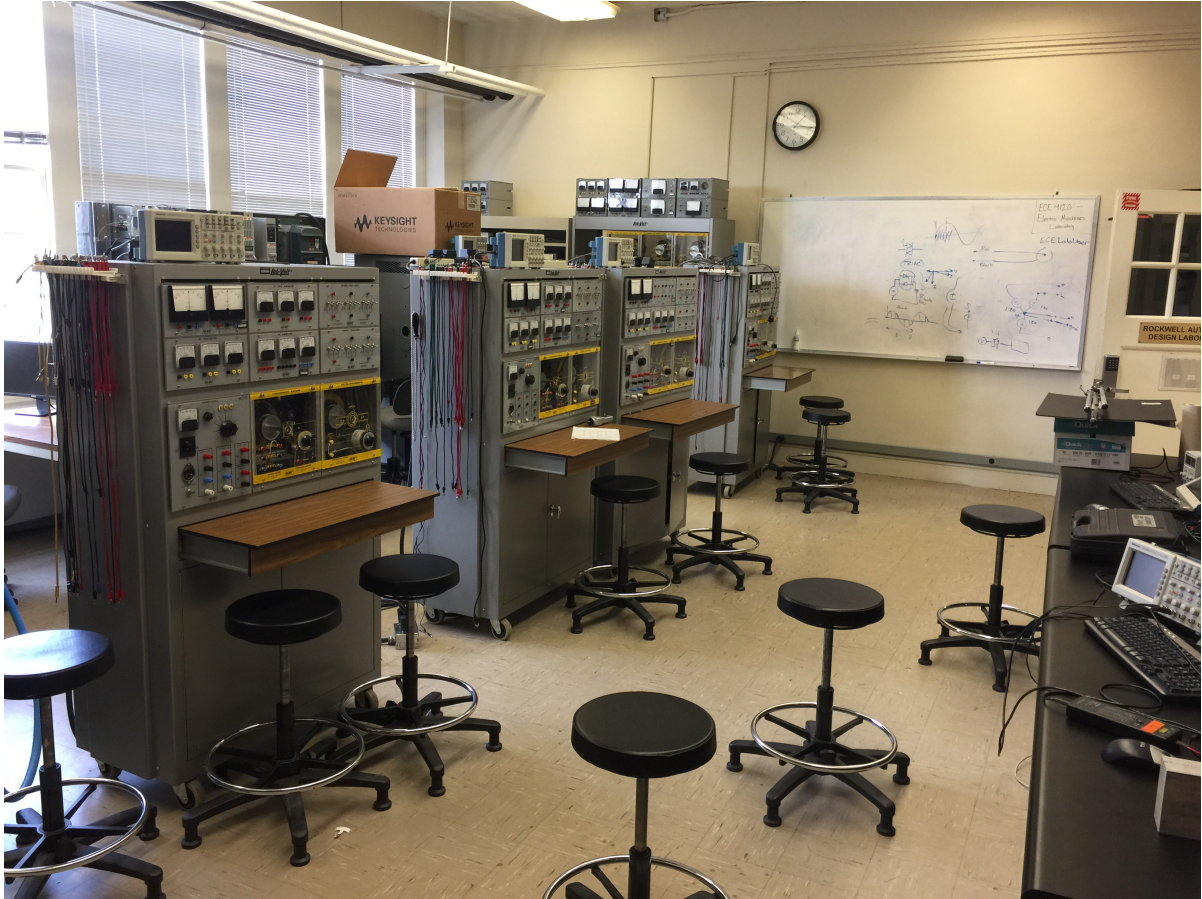


Figure 2.9: One of the rooms searched by participants during the semi-regular gait activity.



Figure 2.10: The original Lego instruction booklet, nearly assembled Lego, and bin containing the final pieces for assembly.

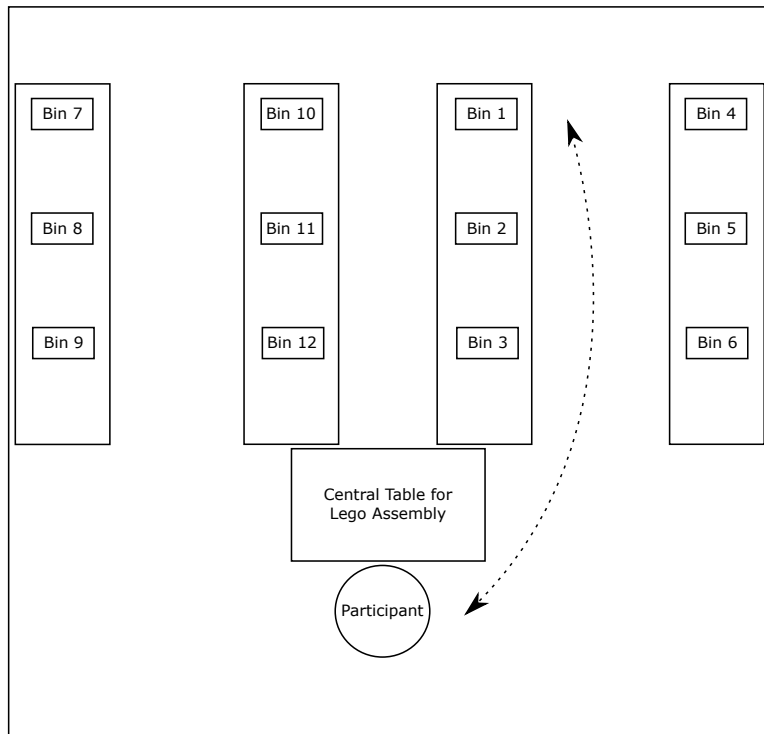


Figure 2.11: A visualization of the process used throughout the unstructured gait activity. Participants construct Legos in a central location while periodically collecting additional bins of lego pieces located around the room.



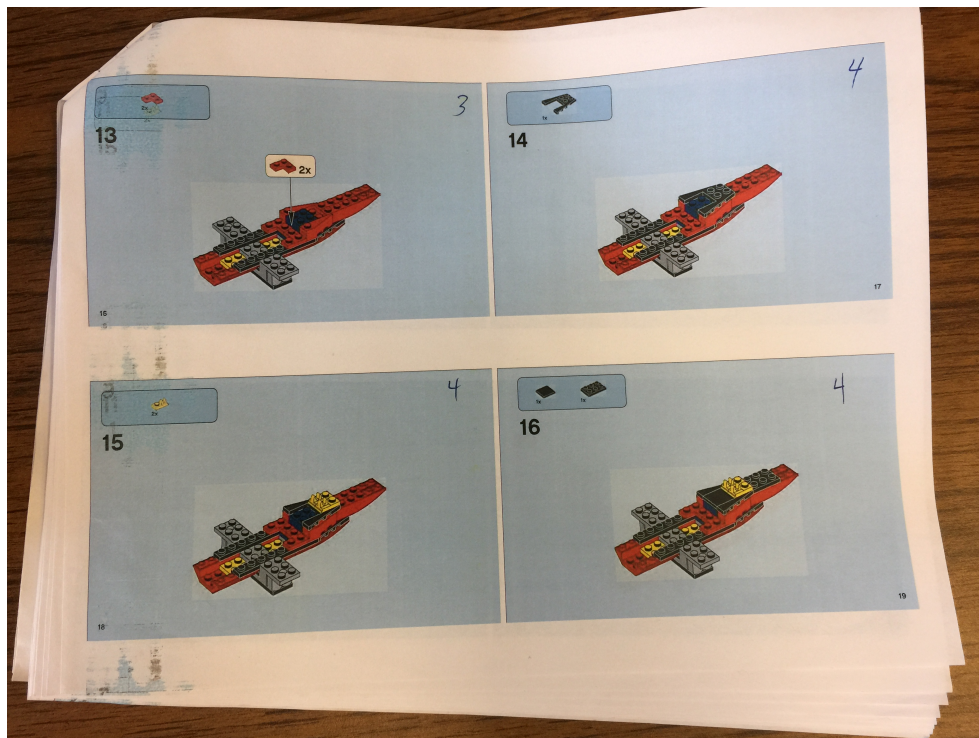


Figure 2.12: A printed copy of the Lego instruction booklet, with handwritten numbers in the top right of each frame indicating which bin contains the pieces necessary for completing the corresponding step.



Figure 2.13: The Fitbit Charge 2 was worn by each participant to assess the accuracy of a commercial device in each activity.

Upon completion of each activity, each participant was told to remain still as the number of steps displayed by the Fitbit were again recorded and the Shimmer3 devices were turned off. Between activities, participants were free to rest as needed. By calculating the difference in steps reported by the Fitbit before and after each recording, the Fitbit RCA (running count accuracy) could be evaluated.

### **2.1.2 Patterns in Regular Gait**

In general, when walking with a regular gait, a repeating motion is present across all parts of the body. The core moves up and down with each step, the arms move in pendulum arcs across two steps, and the legs follow a repeating pattern of motion as well. While each of these locations has a regular motion, differences in how the motion is achieved serve to generate differing accelerometer signals, as shown in Figure 2.14. The signal generated from each sensor location is unique, but each exhibits a clear pattern that can be tied to specific motion patterns in walking.

The wrist signal is generally noisier than the hip and ankle signals. While walking, the arm swings in a pendulum pattern, but there are multiple components to its motion. In addition to the

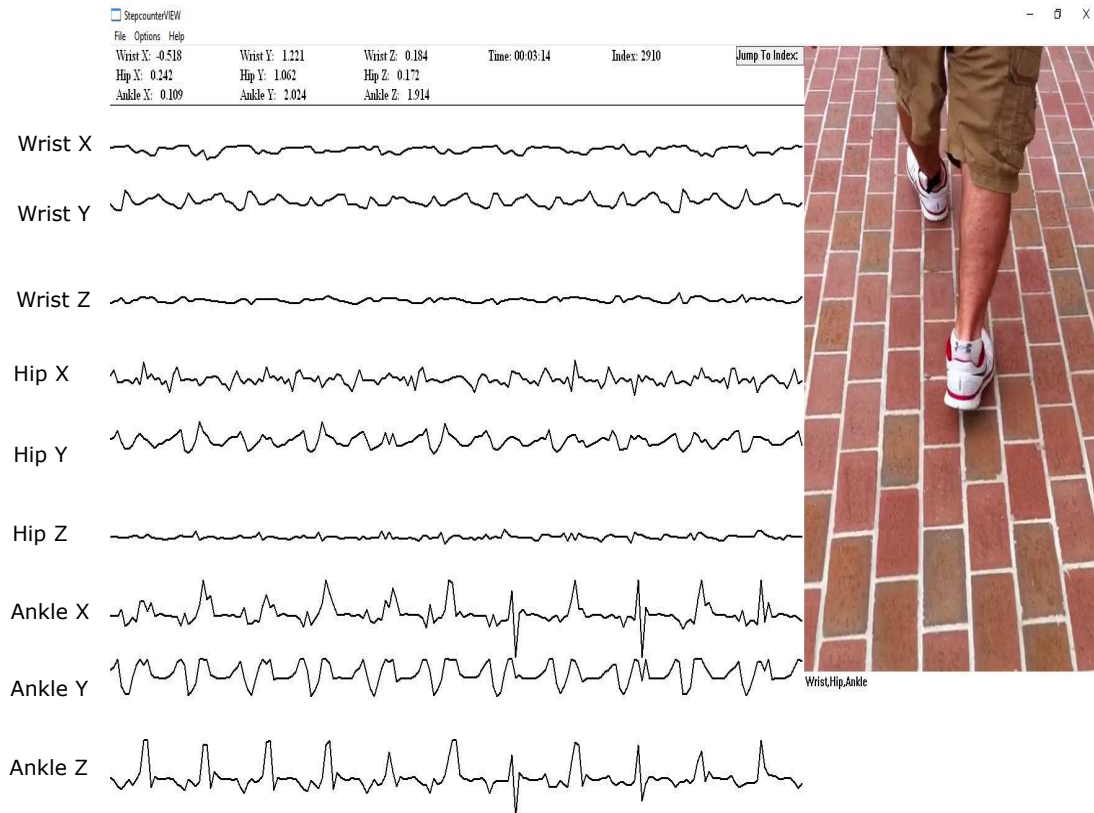


Figure 2.14: Example of acceleration measured for each location and axis. Acceleration peaks related to each step are most visible in the Wrist Y axis and Hip Y axis. For the ankle, the characteristic step pattern is a period of no motion (as the non-instrumented step is moving), followed by two periods of acceleration as the instrumented begins moving and stops moving.

forward and backward swing of the arms, they also move up and down with the core of the body. From an acceleration standpoint, the four largest accelerations generated result from the two points when the arm reaches the apex of its swing and changes direction and the two points when the body switches between upward and downward motions. It is likely that the combination of all of these acceleration sources causes more variability in the magnitude of the acceleration signal in the signal from the wrist sensor.

The signal from the hip worn sensor provides a clearer pattern of continual acceleration changes. Each peak in the magnitude of acceleration signal corresponds to a step as the core moves up and down. This pattern may be slightly affected by the side of the body being instrumented, but this has less of an effect on the signal because the core, as a whole, moves with each step.

The signal generated by the ankle sensor varies greatly depending on which foot is taking the step. For example, if the left ankle is instrumented and a step is taken with the right foot, the instrumented ankle remains relatively still. On the other hand, if the left ankle is instrumented and a step is taken with the left foot, there will be two periods of significant motion. The first as the left foot is lifted off the ground and the second as it is planted. These two motions are characterized by very large accelerations as the foot transitions from rest to motion and from motion to rest. This pattern, from one ankle sensor, can be characterized by a period of rest while the non-instrumented ankle moves, followed by a sharp acceleration as the foot is picked up and a sharp deceleration as the foot is planted on the ground.

Because each signal varies depending on axis, an algorithm developed to use the Y-axis, for example, would not work if the sensor was rotated or positioned differently. In order to allow algorithms to operate independent of orientation, it is common for algorithms to begin by calculating the magnitude of acceleration across all three axes and using this measure as the basis for the algorithm. This is typically accomplished by applying Equation 2.1 to the three axes of data provided by one accelerometer,

$$acc_m(t) = \sqrt{acc_x(t)^2 + acc_y(t)^2 + acc_z(t)^2} \quad (2.1)$$

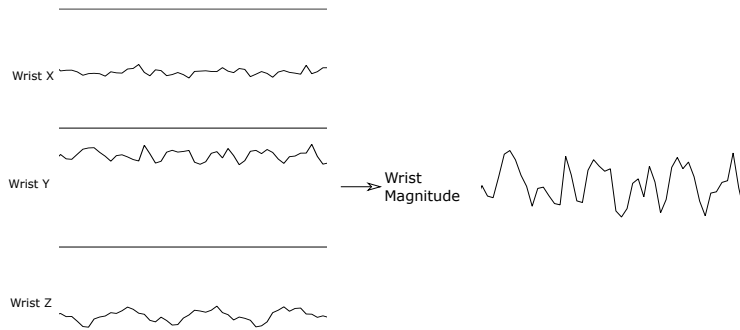
where  $acc_m(t)$  is the magnitude of the acceleration signal at time  $t$ , and  $acc_x(t)$ ,  $acc_y(t)$ , and  $acc_z(t)$  are the acceleration on the x, y, and z axes, respectively, at time  $t$ . An example of how this process affects the signal for the Wrist, Hip, and Ankle are shown in Figures 2.15a - 2.15c. Each of the

three signals shown has a different appearance, however, when examined together, as in Figure 2.16, it becomes clear that there are significant similarities between sensor positions. In the figure, the participant is walking with a regular gait, and their steps are identified with either a red solid line for steps taken with the right foot or a blue solid line for steps taken with the left foot. While the signals do have differences, there is clearly one peak associated with one step for each body position. Because of this consistency across sensor positions, it is possible to use one step detection algorithm for all body positions.

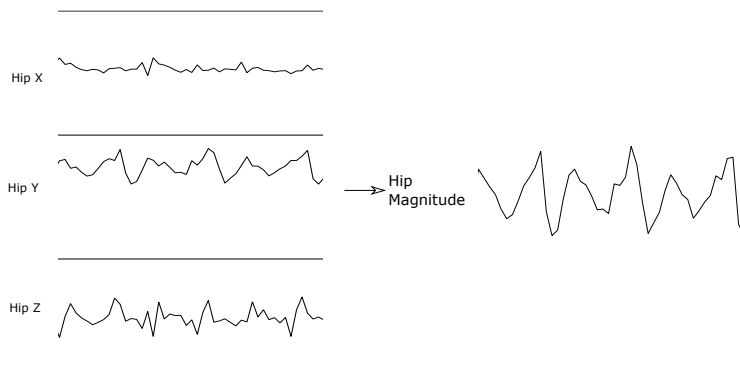
### 2.1.3 Challenges in natural gaits

When free living is considered, the accelerometer signal generated by motion can differ significantly from the regular gait accelerometer signal. In this paper, we consider three different walking patterns typical in daily life: exercise, moving around a building, and moving around a room. Exercise typically involves a period of repetitive motion such as walking on a treadmill, walking around a track, or walking down a road. Moving around a building typically involves multiple periods of shorter duration movement, as when walking from one room to another, interspersed with periods of moving about each room. This could also include periods of walking up or down stairs. Examples include walking to a meeting from an office or walking to a classroom for a lecture. Moving around a room typically involves very brief periods of walking interspersed with periods of time in which significant motion occurs but no steps are taken. Examples of this include cooking, folding and putting away laundry, or cleaning the house. We refer to these walking patterns as regular (exercise), semi-regular (moving around a building), and unstructured (moving around a room).

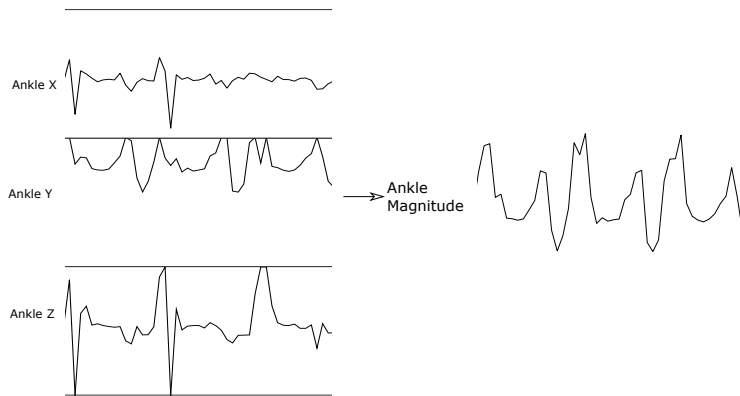
Examples of the accelerometer signals generated from regular, semi-regular, and unstructured gaits can be seen in Figure 3.8. These gait types demonstrate the challenges in developing an algorithm for walking motions present in natural living, as compared to only developing for regular gait. In Figure 2.17a, the peaks corresponding to steps taken with the instrumented ankle are identified with arrows. It is clear that each of these peaks is easily detectable with a relatively wide range of thresholds. In Figure 2.17b, the participant transitions from walking down a hallway to climbing stairs. Even though peaks are discernible for all steps taken in this example, it is clear that the accelerometer pattern is different, and thresholds used would have to adapt based on which part of the signal is analyzed. The greatest challenge for detecting steps can be seen in the unstructured



(a) Finding the magnitude of the x, y, and z axes for the wrist.



(b) Finding the magnitude of the x, y, and z axes for the hip.



(c) Finding the magnitude of the x, y, and z axes for the ankle.

Figure 2.15: Examples of finding the magnitude of an accelerometer signal divided across a 3-axis accelerometer.

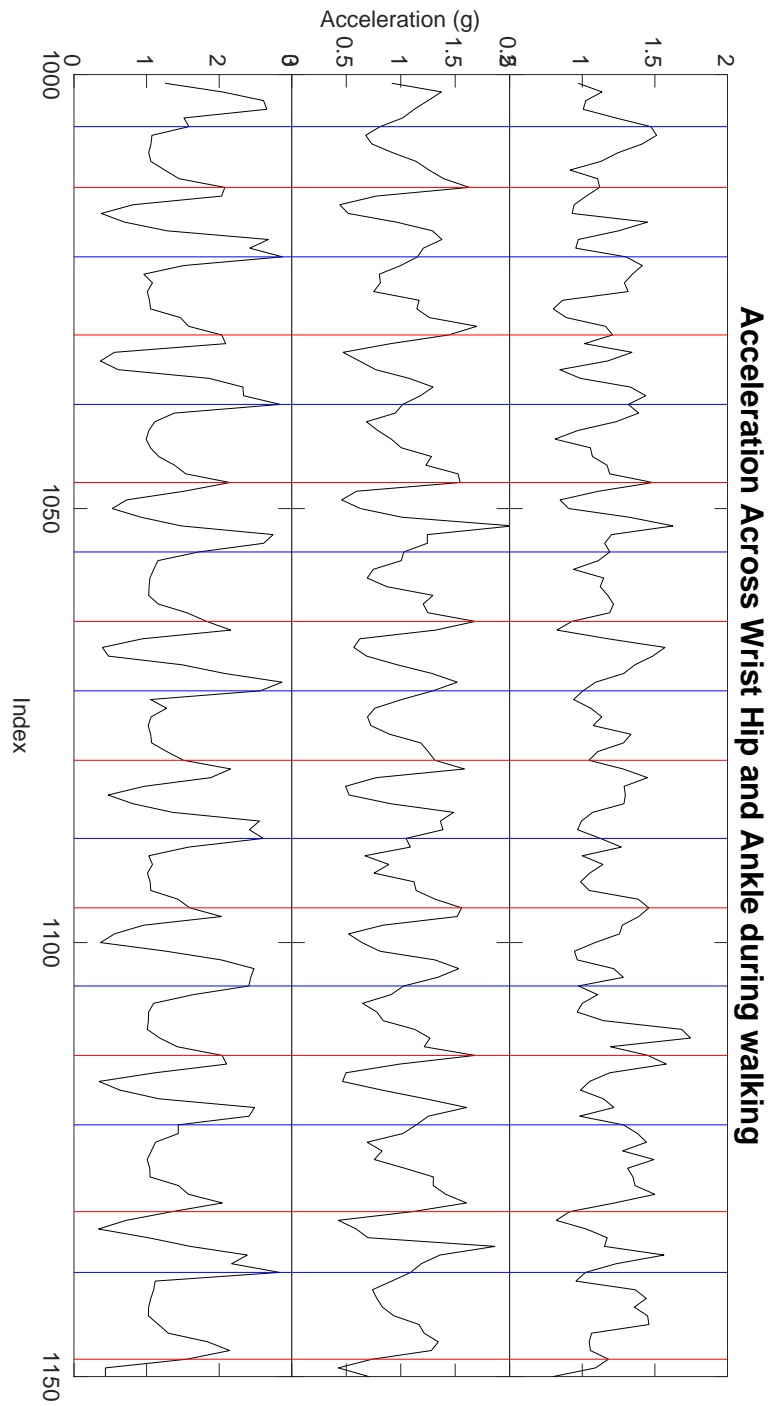


Figure 2.16: Example of different acceleration signals for the wrist (top), hip (middle), and ankle (bottom) during regular walking.

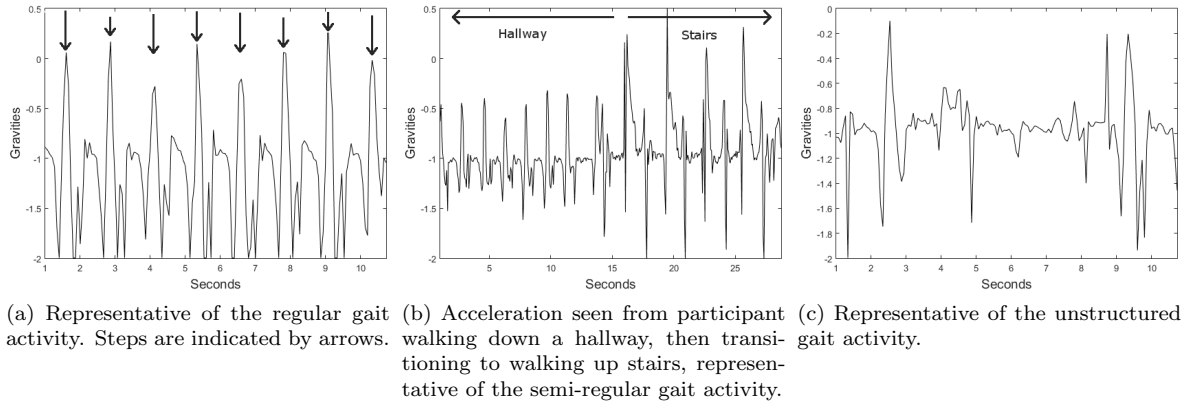


Figure 2.17: Examples of accelerometer data collected from the three different activities performed.

gait example shown in Figure 2.17c. In this example, it is unclear from the signal when steps are taken. More advanced algorithms are needed to perform well on this signal.

Prior algorithms and datasets have all been developed with an emphasis on regular gait. This includes walking, running, or climbing stairs at gaits from slow walking to running, as long as the participant is in regular motion. The cyclical nature of the signal produced has been used to improve algorithms in a number of ways. One example of this can be commonly seen in peak detection algorithms. After local minima or maxima are detected, these algorithms frequently require that the peaks occur within a specific range of times before or after another peak. The peaks are only treated as steps once a set number of consecutive peaks meet these time requirements. This type of classification requirement only works well when regular gait is considered. When examining steps in free living, the time between steps can vary greatly, and the number of consecutive steps can be very small. If the consecutive peak requirements are applied, a large number of peaks may be excluded when they actually represent a step being taken.

### 2.1.4 Ground Truthing

Once the data was collected for all participants, a custom software tool was developed for data visualization and the identification of ground truth steps. The tool developed to accomplish this task is shown in Figure 2.18. The tool displays the accelerometer signals and video simultaneously and allows the reviewer to mark steps, including information about which foot took the step. There are nine accelerometer signals shown. The top signal is the x-axis from the Shimmer3 device worn



on the non-dominant wrist. The second and third rows are the y and z axes from the wrist device. This pattern is repeated for the device worn on the hip and the device worn on the ankle, providing the 9 total signals shown in the figure. The green rectangle is the indicator of the current index being examined. Since the Shimmer3 devices were set to record at 15 Hz and the video was recorded at 30fps, signals could be examined and steps identified with a 67 millisecond level of precision. Reviewers can add or remove steps, and the steps are displayed as vertical lines, with solid red lines for steps taken with the right foot and dashed blue lines for steps taken with the left foot. Identifying the correct foot is of particular importance since only one wrist and one ankle are instrumented, so the accelerometer signals may look significantly different depending on which foot is taking the step. In addition to the graphical representation of the acceleration for the x, y, and z axes for each sensor shown in the plot, the tool also displays the numerical values for the raw accelerometer data (in gravities) at the very top of the image. In addition, the time in the recording and current index being examined are displayed. Reviewers are able to move forward and backward and play and pause. In addition, reviewers can jump to an index through use of a dialog box. Through menu selections, reviewers can select to display only one sensor at a time. Finally, the reviewer can zoom in or out on the data in order to observe finer detail or examine how accelerometer patterns change over wider swaths of time.

In the segment of data shown in Figure 2.18, the participant is walking during the regular gait activity. Of the three wrist axes, the greatest motion is seen on the X and Y axes. This is because the wrist worn sensor captures the rhythmic oscillations present as the arm swings back and forth (the up and down motion of the body, and the forward and backward motion of the arm as it swings in the direction of walking provide the largest portions of motion). Because there is relatively little motion in the wrist on the axis perpendicular to the direction of motion (this would be seen if the arm moved inward, towards the hip or outwards away from the hip), the Z axis for the wrist motion in the figure is relatively inactive. Similarly, the hip sees vertical motion as the body moves up and down with each step as well as motion in the direction of walking. There is little motion perpendicular to the direction of walking. Unlike the wrist, in which both the X and Y axes have significant accelerations for each step, the hip's motion is more concentrated on a single axis, the Y axis, because the greatest accelerations are seen in the upward and downward motion of the body (the forward acceleration is relatively constant, compared to the wrist, because the additional accelerations of moving the arm forward and backward are not present). The ankle

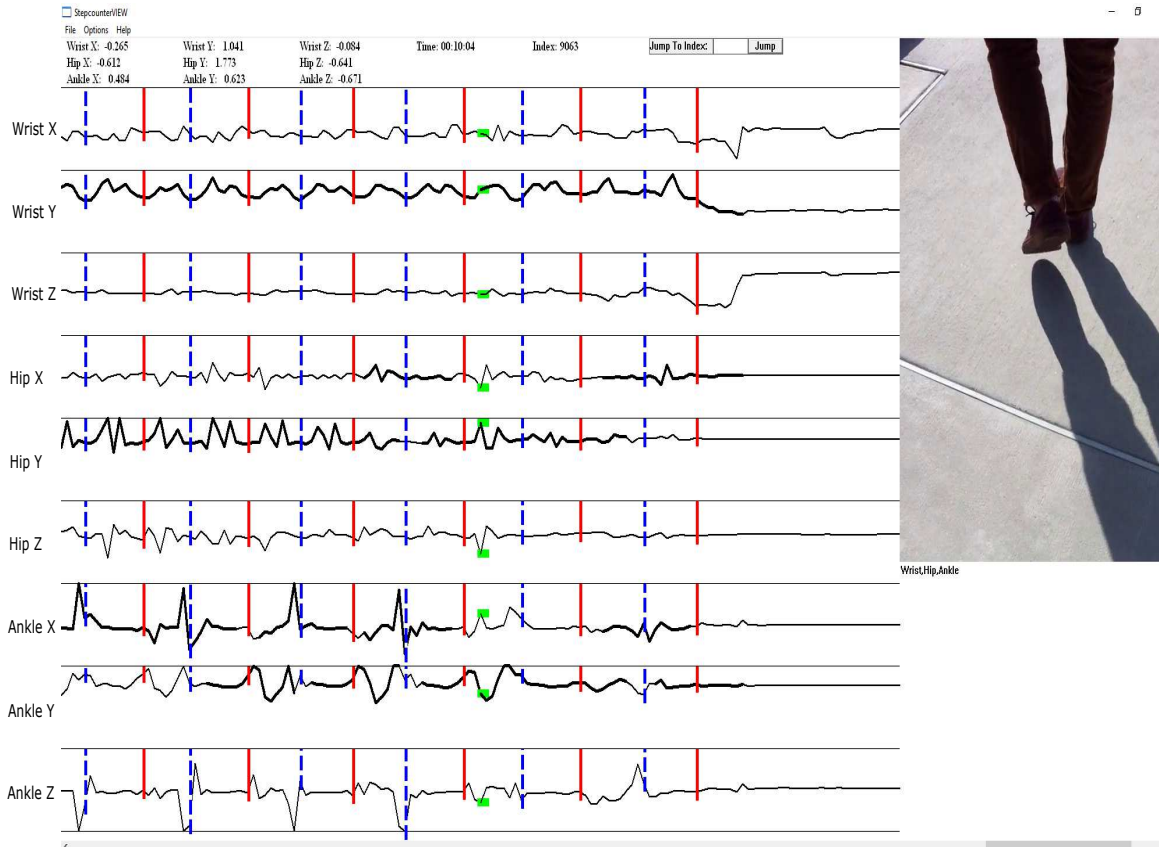


Figure 2.18: The tool developed for visualizing the accelerometer data and identifying ground truth steps. The 9 axes of accelerometer signal (x, y, and z axis for wrist, hip, and ankle) are displayed on the left, and the video recording is displayed in the top right. Steps taken with the right foot are marked with a solid red line and steps taken with the left foot are marked with a dashed blue line.

exhibits a unique pattern, compared to the wrist and hip. The ankle follows generally displays two moments of acceleration followed by a relatively long period with no acceleration whatsoever. Since only the left ankle is being examined, there is essentially zero acceleration in the time period in which the left foot is planted on the ground as the right foot takes a step. There is then substantial acceleration as the left foot is picked up and transitions to moving forward. A second period of acceleration occurs when the foot transitions from moving forward to stopping and planting on the ground.

The data was reviewed by this author as well as Sriram Madhivanan. Twenty participants were annotated by this author, and ten participants were annotated by Sriram, with additional review and revision by this author to ensure consistency across participants. The manual annotation process took approximately 4 hours per participant for a total of 120 hours of annotation. A total of 60,853 steps were identified and marked.

### **2.1.5 Metrics**

There is a tendency for performance or effectiveness to be evaluated based on a single number. For example, academic journals are often reduced to impact factor. Cars are reduced to a reliability score. Windows reduces a computer's effectiveness to a Windows Experience Index. These values do not provide enough information to fully evaluate effectiveness. A similar phenomenon can be seen in pedometer evaluation. Running Count Accuracy (RCA), or a long term measure of the number of steps detected compared to the actual number of steps taken, is the only evaluation metric used in the vast majority of pedometer studies. While RCA can provide some useful information about the pedometer algorithm's effectiveness, multiple metrics are needed to fully understand where algorithms are accurate or inaccurate. This work considers two metrics for evaluation, run count accuracy (RCA) and step detection accuracy (SDA).

While RCA can be useful for providing a rough estimation of a pedometer algorithm's effectiveness, it does not provide an accurate assessment if the number of false positives and false negatives produced by an algorithm are significant, but equal in frequency. This could be a problem, for example, if in the Lego building activity a pedometer algorithm treats many of the motions used when building the Lego as steps, but does not identify the steps taken while the participant moves to pick up additional pieces. While the overall running count accuracy may be similar, the algorithm was not effective at detecting steps, and this problem may only manifest itself in another activity.

The RCA value is the value most commonly reported in prior works, however, so it is reported as an evaluation metric in this work as well. The equation used to calculate RCA is shown in Equation 2.2. This evaluation metric does not always provide an accuracy between 0 and 1. If the number of detected steps is greater than the number of ground truth steps, the resulting value will be greater than one. Thus, an accuracy of 1 is ideal, an accuracy less than 1 indicates that an algorithm under-detects steps, and an accuracy greater than 1 indicates that an algorithm over detects steps.

$$RCA = \frac{\#DetectedSteps}{\#GroundTruthSteps} \quad (2.2)$$

Step detection accuracy (SDA) provides a more detailed assessment of pedometer accuracy. SDA examines the ability of an algorithm to specifically identify individual steps. This allows evaluation through classification of steps as true positives, false positives, and false negatives. True positives are identified when a step detected by an algorithm occurs within 1/2 second of a ground truth step. These two detections are paired and excluded from further evaluation (eg: if the next step detected by the pedometer algorithm is within 1/2 second the ground truth step paired with the previous detection, the ground truth step cannot be paired again.) Any step detected by the algorithm which cannot be paired with a ground truth step is considered a false positive. At the end of this process, any ground truth step not paired with a step detected by the algorithm is recorded as a false negative. True negatives are not considered because each algorithm would report high accuracy due to the high sample rate relative to the frequency of steps. For example, in our regular gait activity, which has the highest frequency of steps, the sample rate is 15 Hz and the frequency of steps is less than 2 Hz.

Because true negatives are not considered, there are two metrics which can serve to evaluate accuracy. Positive predictive value (PPV), also known as precision, is calculated according to Equation 2.3 and provides a measure of how likely a step detection made by the algorithm is to indicate that a step is actually taken. Sensitivity, also referred to as recall, is calculated according to Equation 2.4 and provides a measure of how likely an algorithm is to detect a step when one is taken. An algorithm should provide good results for each of these measures in order to be effective, and the F1 score is a standard measure which evaluates accuracy based on a combination of PPV and Sensitivity through Equation 2.5. The F1 score is similar to an average of PPV and sensitivity, but compared to a raw average, the score is decreased when the two metrics are significantly different.

For example, a PPV of 0.1 and a sensitivity of 0.9 would average to 0.5, but the F1 score is actually .18. On the other hand, a PPV of 0.5 and a sensitivity of 0.5 would average to 0.5 and also have an F1 score of 0.5. Essentially, this penalizes algorithms which lean too heavily toward extremes: counting all data points as steps or no data points as steps.

$$PPV = \frac{TruePositives}{TruePositives + FalsePositives} \quad (2.3)$$

$$Sensitivity = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (2.4)$$

$$F1 = \frac{2xPPVxSensitivity}{PPV + Sensitivity} \quad (2.5)$$

Evaluations using these two metrics allow for a more thorough understanding of when, where, and how pedometer algorithms fail. RCA is a limited metric because regardless of how it is calculated, if the number of false positives and false negatives are similar, the reported RCA will always be near 1. In an extreme example, imagine a participant wears a pedometer throughout their full day, but the pedometer does not detect any steps until they lie down to go to sleep. This hypothetical pedometer then detects this action and it triggers 7,000 steps in the algorithm. If the participant walked 7,000 steps in the day and don't look at the pedometer until after they have lain down, they would claim that the pedometer is accurate, and according to RCA measures, they would be correct. In reality, however, this is an extremely ineffective algorithm. In the same situation, SDA would correctly indicate that the algorithm is extremely inaccurate because there would be 7,000 false positives, 7,000 false negatives and 0 true positives. RCA does, however, provide some useful information regarding algorithm performance. Based on our method of calculating RCA, if the RCA is significantly greater than 1, it can quickly be learned that the algorithm is, overall, significantly overestimating steps, and a value significantly less than 1 indicates a significant underestimation of steps. A value near 1 does not indicate accuracy, necessarily, but does indicate that false positives and false negatives are roughly equal. SDA, represented by an F1 score, provides information about how effective the algorithm is at detecting individual steps, but, as a single measure, it does not provide information about how much an algorithm is over or under estimating the long term step count. Based on this information, this work considers SDA as a more effective measure of pedometer

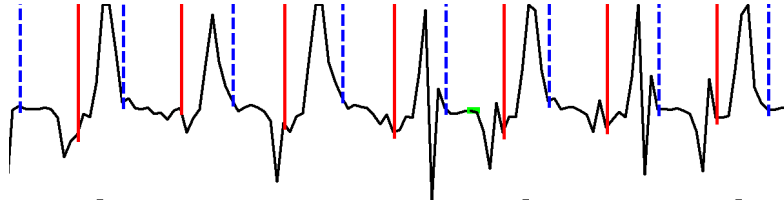


Figure 2.19: An example of ground truth steps identified in the regular gait activity. The peak in acceleration between each right step (red solid line) and left step (blue dashed line) is indicative of the forward motion of the left leg.

algorithm accuracy, while using RCA to interpret how the inaccuracies present in the algorithm affect the long term step count.

### 2.1.6 Definition of a Step

The annotation process revealed that steps could be categorized into two distinct groups, which we call steps and shifts. In the regular gait activity, because the motion of walking is repetitive and uninterrupted, each step taken can be clearly counted and evaluated. Virtually all steps in the regular gait activity can be defined by the inclusion of three elements: 1) the foot moves, 2) the body weight shifts in the direction of the foot movement, and 3) the action takes place within a repeating pattern. An activity meeting these criteria is called a step. Steps typically provide a clear accelerometer signal for identification, as shown in Figure 2.19. In this figure, displaying acceleration captured from the ankle accelerometer, it can be seen that between every right step (red solid line) and left step (blue dashed line), there is a clear peak in the acceleration signal as the left leg moves forward.

While steps do occur within the semi-regular and unstructured gaits, a significant number of motions occur which contain a foot movement, may or may not include a weight shift, and are not within a repeating pattern. We call this second type of motion a shift. Shifts are identified most frequently when one of three different movements occur: 1) the first or last stride in a sequence of steps, 2) shuffles (weight shifts accompanied by small foot movements), and 3) pivots (one foot rotates while the other is planted). Examples of each of these subcategory of shift are displayed in Figure 2.20. The first two rows in the figure demonstrate first and last steps taken and demonstrate a smaller magnitude of acceleration compared to steps taken while in stride. The third row demonstrates a pivot step, which occurs as the participant changes direction. The fourth row demonstrates a shuffle,

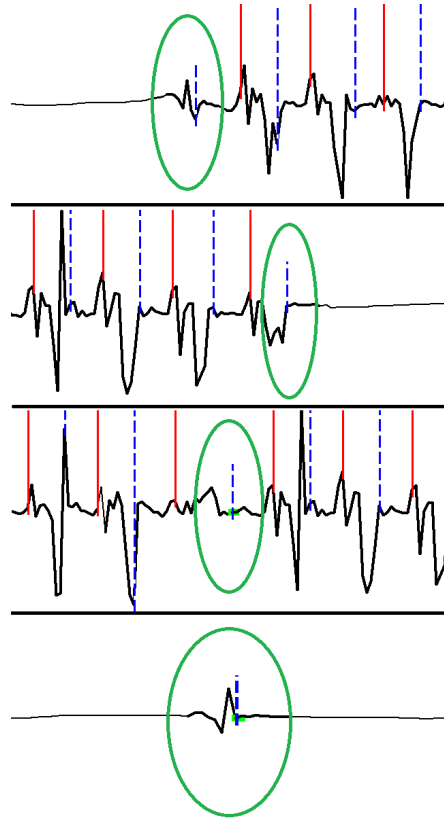


Figure 2.20: An example of the accelerometer signal associated with four different types of shifts. The first two rows provide examples of first and last steps being taken. The third row demonstrates a pivot. The fourth row demonstrates a shuffle step.

in which the participant moved their foot and shifted their weight while building the Lego toy in the unstructured gait activity. Shifts have a different appearance in the accelerometer signal and we expect shifts to produce a different level of energy expenditure when compared to traditional steps.

## 2.2 Results

This chapter discusses two types of analysis performed on the dataset collected, providing results regarding the prevalence of steps and shifts and the accuracy of a commercial pedometer for each activity. This analysis was made possible through manual annotation of the data, a process which took approximately 4 hours per participant, requiring a total of 120 hours of annotation. Through this process, a total of 60,853 steps were identified and recorded. This analysis allows the dataset to be used for detailed pedometer evaluation.

Activity	Steps	Shifts
Regular	31,447 (99.7%)	88 (0.03%)
Semi-regular	19,919 (89.6%)	2,317 (10.4%)
Unstructured	5,263 (74.3%)	1,819 (25.7%)
Overall	56,629 (93.1%)	4,224 (6.9%)

Table 2.1: Summary of steps and shifts manually recorded through each of the three activities.

### 2.2.1 Steps and Shifts

One critical finding revealed through the evaluation process was the presence of two distinct categories of steps, which we call steps and shifts. We define steps as motions which fulfill three requirements: 1) the foot moves, 2) the body weight shifts in the direction of the foot movement, and 3) the action takes place within a repeating pattern. In regular gait activities, the walking motion is typically repetitive and uninterrupted, causing almost all steps taken to fit all the pieces of this definition of a step. Prior work has focused heavily on regular gait analysis, which may explain why the second type of step we consider has not been discussed. This second type of step is called a shift, which can be similarly defined with three requirements: 1) the foot moves, 2) the body weight may or may not shift in the direction of foot movement, and 3) the action does not take place within a repeating pattern. While this type of step is extremely rare in regular gait, they do occur with significant frequency in semi-regular and irregular gaits. These shifts are identified most frequently as one of three distinct types of motion: 1) the first or last stride in a sequence of steps, 2) shuffles (weight shifts accompanied by small foot movements), and 3) pivots (one foot rotates while the other is planted). Shifts have a different appearance in the accelerometer signal and we would expect shifts to produce a different level of energy expenditure when compared to traditional steps.

The proportion of steps and shifts present in each of the three gait types is shown in Table 2.1. Within the experiment, each activity was designed to take approximately 10 minutes to complete. Because activities were standardized on time rather than number of steps, the step count for each activity varies significantly. An average of 1,050 steps were taken in the regular gait activity, 667 in the semi-regular gait activity, and 175 in the unstructured gait activity. The vast majority of recorded motions in the regular gait activity were steps (99.7%). The shifts that do occur in regular gait are primarily composed of the starting and stopping steps taken by each participant.

In the semi-regular gait activity, the number of shifts increased to 10.4% of the total steps



taken. These shifts included a larger number of starting and stopping steps taken as participants stopped to open doors throughout the activity. In addition, while hunting for the hidden items throughout rooms, participants started, stopped, shifted their weight, and pivoted. The steps taken while walking down hallways between rooms and up and down staircases still made up the majority of motion, however, and these activities were primarily composed of steps.

In the unstructured activity, the percentage of shifts increased again, where 25.7% of the motions recorded were shifts. In this activity, participants spent most of their time building the Lego toy at a central location, periodically interrupted with brief periods of walking as participants collected additional pieces from bins located around the room. While building the Lego, participants shifted their weight, causing some shifts. The starting and stopping steps while walking to get additional Lego pieces around the room were the source of many of the shifts. Another significant portion of the shifts resulted from pivots performed when participants picked up a bin of pieces and turned to return and continue building the Lego.

Overall, as the motions performed in the activities became less repetitious, the percentage of motions classified as shifts increased. In regular gait, the shifts were extremely rare, but in unstructured gait, the shifts composed over 1/4 of the total motions recorded. This implies that shifts may not be considered at all in studies examining only regular gait because shifts are extremely rare and could easily be treated as steps. When examining semi-regular and unstructured gaits, however, the shifts increase to a point that they cannot be ignored. Future work in pedometer evaluation for these gaits, which are present in natural walking, must design and develop a method for evaluating shifts.

### **2.2.2 Fitbit Accuracy**

The three gait types we examined can also provide insight when evaluating commercial pedometer performance. To this author’s knowledge, commercial pedometer algorithms are proprietary, and therefore cannot be reimplemented on a custom dataset. Because of this limitation, individual step detection accuracy (SDA) cannot be evaluated in proprietary commercial pedometer algorithms. It is possible, however, to compare running count accuracy (RCA) provided by commercial pedometers with the ground truth step count determined for each activity. In the collection of our dataset, each participant wore a Fitbit directly adjacent to the Shimmer3 device on their non-dominant wrist. The step count displayed on the Fitbit device was recorded before and after

Activity	Steps	Fitbit (RCA)
Regular	31,528	30,318 (96.2%)
Semi-regular	22,225	19,679 (89.7%)
Unstructured	7,081	4,552 (64.3%)
Overall	60,834	54,815 (90.1%)

Table 2.2: Comparison of ground truth steps with number of steps reported by Fitbit.

each portion of the experiment, providing a Fitbit step count for the regular, semi-regular, and unstructured gait activities independently.

Fitbit accuracy was first evaluated against all recorded motions, including both steps and shifts, with results summarized in Table 2.2. The Fitbit underestimated the overall number of steps taken for each of the three gait types. The Fitbit demonstrated the highest accuracy for the regular gait activity (96.2%). This result makes sense if it is assumed that the Fitbit algorithm is trained on regular gait, as many published pedometer algorithms are. The Fitbit accuracy decreases to 89.7% in the semi-regular gait activity. If it is assumed that the detection of shifts is challenging for the commercial pedometer, this result make sense. With an increase of roughly 10% in the percentage of steps which are considered shifts, an accuracy drop of 6.5% is within a reasonable range. The Fitbit’s accuracy dropped to 64.3% in the unstructured gait activity. Since shifts compose roughly 25% of steps in this activity, it is somewhat surprising that the accuracy drop from the regular gait activity to the unstructured gait activity is 31.9%. This drop could indicate that a minimum number of consecutive steps is required for the Fitbit to count steps, and because the portions of the activity that include walking are relatively brief (3 - 10 steps), some of these walking segments do not trigger Fitbit’s step detection algorithm. This theory is speculation because Fitbit’s step detection algorithm is proprietary, but it does raise some interesting questions regarding the algorithm’s implementation. Overall, because there are more steps in the regular and semi-regular gait activities than the unstructured gait activity, the Fitbit’s accuracy was 90.1%.

The Fitbit’s accuracy was also evaluated against only steps, removing consideration of shifts, as summarized in Table 2.3. Because the Fitbit underestimated steps in each of the activities, removing shifts from consideration increases accuracy for all activities. In the regular gait activity, there were only a small number of shifts, so the accuracy of the Fitbit is not affected significantly. In the semi-regular gait activity, removing shifts actually causes the Fitbit to slightly overestimate the number of steps taken. In addition, the accuracy of the Fitbit for steps only in the semi-regular

Activity	Steps without shifts	Fitbit (RCA)
Regular	31,435	30,318 (96.4%)
Semi-regular	19,679	19,945 (101.4%)
Unstructured	5,148	4,552 (88.4%)
Overall	56,262	54,815 (97.4%)

Table 2.3: Comparison of ground truth steps, excluding shifts, with the number of steps reported by Fitbit.

gait activity results in the highest accuracy across all activities and conditions, with an accuracy of 98.7%. The accuracy in the unstructured gait activity increases from 64.3% to 88.4%, but the lowest accuracy is still reported for this activity, even when shifts are removed for comparison. Excluding shifts from the evaluation process increases the accuracy of the Fitbit from 90.1% to 97.4%. The Fitbit still underestimates the total number of steps.

## 2.3 Conclusions and Future Work

This chapter described the collection and analysis of a new dataset designed for pedometer algorithm evaluation. The dataset is intended to ultimately serve as a publicly available resource for evaluation of step detection algorithms. In order to facilitate this, accelerometer and gyroscope data were collected from 30 participants performing 3 activities each. The data was collected through use of three Shimmer3 devices located on the wrist, hip, and ankle of each participant. The activities performed in the experiment were designed to elicit regular, semi-regular, and unstructured gaits. These three gait types are representative of three common daily activities including exercising, walking around a building, and walking around a room. The steps were manually identified through use of video review, and a second class of step, called a shift, was identified. Shifts are steps which do not occur within a repeating pattern and generally fall in to three categories; pivots, weight shifts, and starting or ending steps. The presence of these shifts increases as activity becomes less structured. In addition, a commercial pedometer, the Fitbit Charge 2, was worn and its reported steps compared to the ground truth. In each activity, the Fitbit underestimated the number of steps taken, with significant underestimation during the unstructured gait activity.

There are significant opportunities for future improvements and expansions to this dataset. Expanding the dataset could include additional sensors worn or placed in additional locations; for example, in a backpack, a hand bag, a pocket, or being held in the participant’s hand. For future

data collection, the sample rate for the Shimmer3 devices could be adjusted to a higher sample rate in order to allow for finer analysis of the frequency domain signal for step detection. The dataset can be expanded to include a wider variety of activities, particularly activities which are challenging for commercial pedometers. Some examples include driving a car, mowing a lawn, folding laundry, and a number of other activities which have anecdotally been shown to cause inaccuracies in commercial pedometer step counts. In fact, from Fitbit’s support website, it is stated that “When working at a desk, cooking, or doing other arm movements, a device on your wrist can pick up some extra steps if it thinks you’re walking. Many of these situations - such as working or cooking - do include a few steps in-between stationary periods so the device tries to give you credit for those steps” [78]. Our work, specifically from the unstructured gait activity, designed to simulate cooking, indicates that Fitbit may be significantly underestimating these steps. A wider variety of commercial pedometers could be used in future data collection in order to determine if this phenomenon is present across multiple commercial devices.

While commercial pedometer algorithms are generally proprietary, and therefore difficult to re-implement exactly, there are some sources of information regarding these algorithms. For example, according to a Fitbit help topic, the Fitbit algorithm “is designed to look for motion patterns that are most indicative of people walking. The algorithm determines whether a motion’s size is large enough by setting a threshold. If the motion and its subsequent acceleration measurement meet the threshold, the motion will be counted as a step. If the threshold is not met, the motion won’t be counted as a step” [79]. This information is not enough to be certain regarding the specific methods used for step detection, and it certainly does not allow for the algorithm to be reimplemented. The description from Fitbit does seem to indicate that the algorithm is based on some combination of peak and/or threshold detection.

One finding from this work which can be explored in more detail is the presence of shifts. Based on initial evaluation of the dataset, a definition for step and shift were determined. These two actions can be examined in more detail through analysis of the surrounding accelerometer or gyroscope signals. In addition, evaluation of energy expenditure across the two activities could be performed. Based on the signal and energy expenditure analysis, it is possible that the shift activity should be broken in to smaller sub-categories which should be identified separately, including pivots, weight shifts, starting steps, and ending steps. Analyzing the signal of shifts and steps could allow for pedometer algorithms to detect shifts more accurately and, in combination with energy expenditure

analysis, provide a more accurate energy expenditure equation using both steps and shifts as inputs. It is our goal to publish the dataset such that it is freely accessible, so future work will also focus on this process.

## Chapter 3

# Implementation of Prior Pedometer Algorithms

### 3.1 Methods

This chapter describes the process used to collect, classify, and implement prior pedometer algorithms on the dataset developed for algorithm evaluation. Prior algorithms developed for step detection have been categorized, and three previously developed step count algorithms have been reimplemented on the dataset. The three algorithms implemented represent state of the art pedometer algorithms using techniques common across several prior works. Parameters for each of the algorithms were trained and tested on the dataset. This chapter describes the analysis process in detail and describes the details and implementation of the three step count algorithms selected for evaluation.

#### 3.1.1 Step Detection Algorithms

In order to evaluate how pedometer algorithms perform on each gait type, multiple pedometer algorithms were reimplemented on the collected dataset. Over 30 individual pedometer algorithms have been developed and implemented in literature we surveyed since 2004. Since it is impractical to implement every algorithm developed in these papers, particularly since enough detail for reimplementation is not available in many of the publications, we divided algorithms into

Type of algorithms	Citation	Implemented
Peak detection	[18, 19, 48, 49, 50, 53, 56, 59, 62, 64, 65, 68, 80]	[48]
Threshold based	[14, 16, 51, 55, 66, 67, 71, 72, 74]	[67]
Autocorrelation	[54, 61]	[61]
Gyroscope based	[23, 57, 70]	
Image processing	[52]	
Frequency analysis	[73]	

Table 3.1: A summary of prior algorithm methodologies and identification of the algorithms selected for reimplementaion.

broader categories of underlying methodologies. We then implemented one algorithm from three of the categories to represent the most commonly studied pedometer algorithms. The algorithms examined and reimplemented are summarized in Table 3.1. As can be seen in the table, 22 of the 29 algorithms analyzed are based on either peak detection or threshold crossing. Two papers used autocorrelation to detect steps and two used the gyroscope signal to detect steps. While gyroscope data was collected for our dataset, the concepts used for step identification within the two algorithms are similar to the accelerometer peak detection and threshold based algorithms. Gyroscope based algorithms are worth examining in the future, as the signal source is different, but we selected the autocorrelation method for reimplementaion because it operates in an entirely different way and should provide different results. In addition, the gyroscope based algorithms are all performed only for the ankle or foot because the rotational motion detected by a gyroscope is clearest only for the ankle. Because the image processing, chain code, and frequency analysis algorithms were only implemented in one paper each, we placed a lower priority on reimplementing these algorithms.

The largest group of prior algorithms rely on the core concept of peak detection. Peak detection is the process of finding peaks; local minima, maxima, or both. The peak detection algorithms often combine the three acceleration axes (x, y, and z) by finding the magnitude of acceleration and process the resulting signal. It is also very common to apply a low pass filter to smooth out high frequency noise. A variety of filters are used, and Butterworth filters are very common. Alternative methods are used to smooth the signal. For example, Li et. al. use chain code as a method for providing a smoothed accelerometer signal for further analysis [62]. Algorithms using peak detection typically find local minima and/or maxima, treat them as potential steps, and add additional requirements or thresholds in order for the peaks to be treated as actual step detections. The thresholds used to eliminate peaks as potential steps vary, but commonly there are temporal

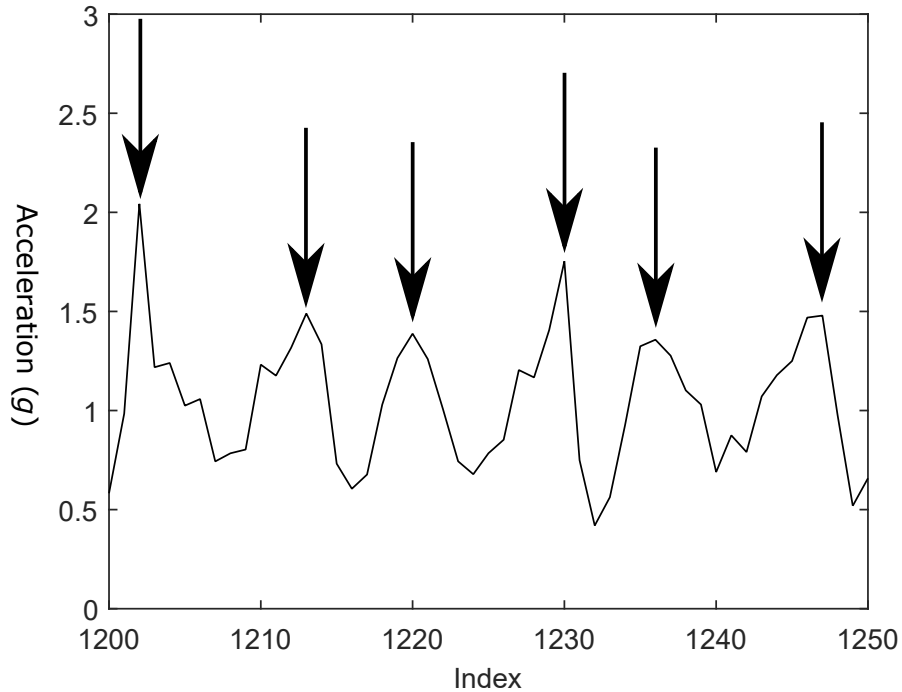


Figure 3.1: Example of peaks detected in an accelerometer signal taken from the hip during regular gait. Peaks are indicated with arrows.

restrictions and magnitude thresholds applied. Temporal restrictions can be applied in two ways. First, in order to detect a peak, maximum or minimum value must be a max or min, not only between its two neighbor values, but also within some span of time. In addition, once maxima and minima are decided, there can be a time restriction from one to the next. For example, for two consecutive peaks to be considered as possible steps, it may be that they are required to occur between 0.5 seconds and 2 seconds of each other. This range is typically determined by the frequency range of human walking (the time between steps when sprinting and walking as slowly as a natural gait will allow). Magnitude thresholds can also be applied by requiring that the magnitude between maxima and minima is large enough in order for these to be considered peaks. With the identification of peaks, and removal of peaks based on thresholds, the peaks that remain are considered steps. An example of peak detection being used to count steps can be seen in the accelerometer sample shown in Figure 3.1. The sample shown is from the hip and was collected during regular gait and peaks which correlate with steps are indicated with arrows.

The second largest group of prior algorithms is categorized as threshold based algorithms.



This category is somewhat broader than the peak detection category, but can be defined by the use of thresholds based on the magnitude of the signal to detect steps. One common way that this is done is through use of “zero” crossings in order to detect steps, where “zero” is determined based on a variable threshold. This threshold is typically dynamic, adjusting to the amount of acceleration in recent samples. For example, the recent maximum and minimum values are found and the “zero” threshold is considered to be the mean of the maximum and minimum values calculated. In order to determine a threshold in real time, the maximum and minimum values from a previous segment of data are used to determine the zero crossing threshold in the present signal. Once the adaptive threshold is calculated, the accelerometer signal is monitored in order to identify when the signal crosses the threshold. When a crossing is detected, the direction of the crossing is also noted, and, depending on the algorithm, only the positive or negative direction crossings are considered as steps. Whether positive or negative crossings are used, they have essentially the same effect: only one out of every two crossings is considered to be a possible step. An example of the adaptive threshold and zero crossing identification can be seen in Figure 3.2. The threshold used for step detection is the blue solid line. This line is the average of the maximum and minimum values from the previous segment of data indicated by the red and green dashed lines, respectively. Whenever the accelerometer signal crosses the blue threshold, this is considered a zero crossing and a step is detected. These crossings are indicated with orange circles.

Step detection through use of autocorrelation is another technique that has been used for step detection. Autocorrelation essentially looks for a repeated pattern in the accelerometer signal and treats each instance of the pattern as a step. This process requires fewer thresholds than peak detection or threshold based algorithms. The algorithm requires essentially three key components for detection. First, the signal must be segmented. Because the frequency with which steps are taken can vary over time, the segmentation process must be adaptive. From one specific data point, the segment of data immediately following the point is compared with the next segment of data. An example of this process is shown in Figure 3.3. The size of these segments is varied, and the size which produces the highest correlation is used. After the correct segment size is determined, if the correlation is above a threshold and the variance in the data is above a threshold, then the two segments are considered steps. The correlation threshold enforces that the similarity between the segments is high enough to indicate a repeating pattern. The threshold on variance is required to avoid treating periods of no motion as steps.

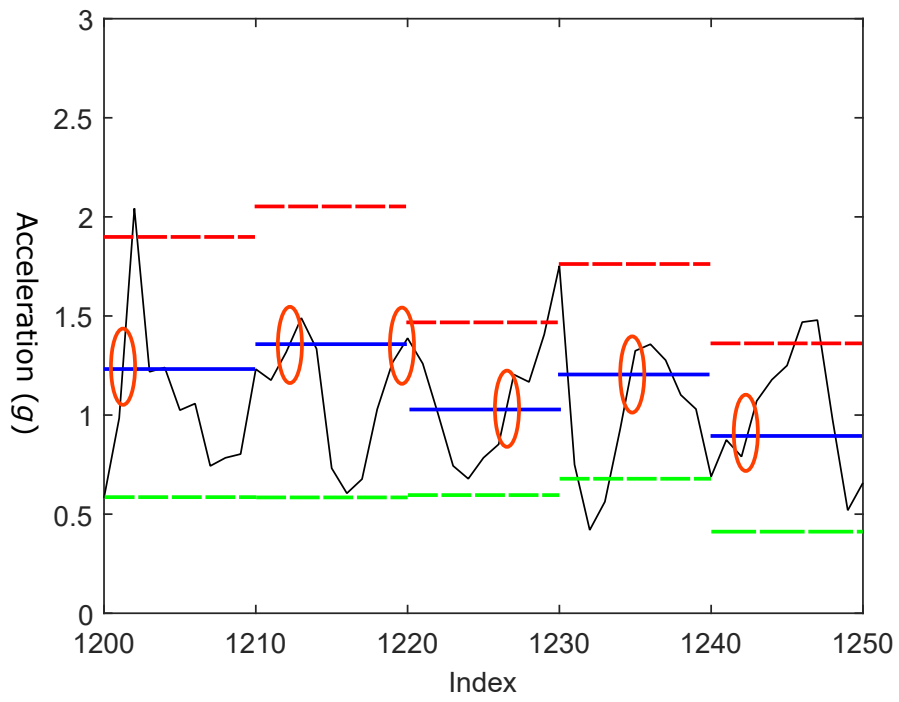


Figure 3.2: Example of zero crossing detection in an accelerometer signal taken from the hip during regular gait. Red and green dashed lines represent adaptive maximum and minimum based on the previous segment of data. The solid blue lines indicate the mean of the maximum and minimum and are used as the threshold for zero crossing detections. Orange circles indicate when the accelerometer signal crosses the threshold, indicating a step.

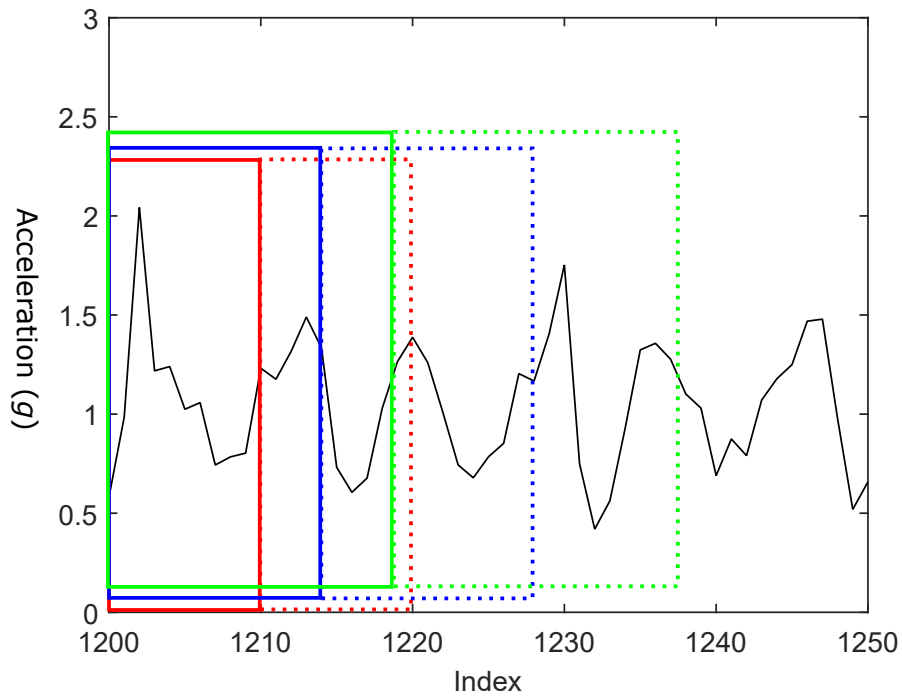


Figure 3.3: An example of obtaining an maximum autocorrelation window and score for index 1200. Starting at index 1200, a range of window sizes are tested for autocorrelation. The smallest window, indicated by a red solid line, is compared against the following window (indicated by a red dashed line). The correlation between these two windows is recorded and compared against correlation scores from a range of window sizes. The highest correlation score and its corresponding window size are recorded and evaluated in the algorithm.

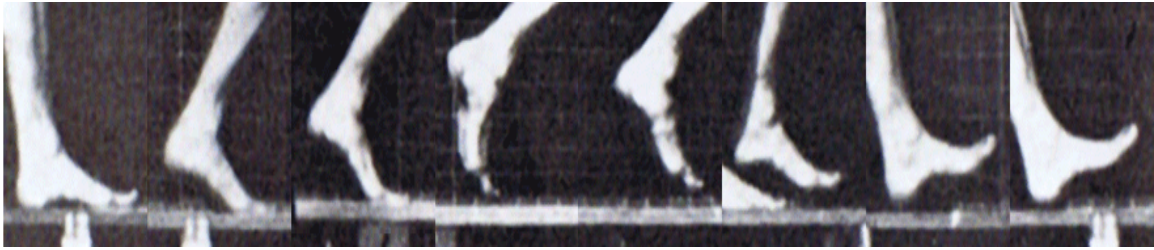


Figure 3.4: Example of how a the foot moves when walking. Gyroscope algorithms are able to measure the rotation in the foot present in this process.

Gyroscope algorithms seek to detect steps based on rotation measurements recorded by a gyroscope. Gyroscopes return angular velocity measurements, as opposed to linear acceleration measures reported by accelerometers. This emphasizes rotational motions over linear motions. In walking, this may cause problems for sensors placed on the core of the body. The hip and chest do not have a significant amount of rotational motion during walking. Wrists can have rotational motion, particularly with fast paced walking, jogging, and running, but when walking at a normal or slow pace, there is relatively little rotation when compared to linear acceleration. The best location for collecting rotational motion is probably the ankle, and in each of the papers using gyroscopes to detect steps, a sensor was located on the foot or ankle [23, 57, 70]. This is because the foot has significant rotation when it is planted and lifted in the walking process, as shown in Figure 3.4. It is possible that alternative locations could be used for gyroscope step detection. Susi et. al. detected steps with based on detecting peaks in a gyroscope signal collected from the foot but also performed step detection using a gyroscope signal from the arm [57]. A standardized dataset is necessary for evaluating the effectiveness of gyroscope step detection algorithms at various sensor locations.

Ozcan et. al. used a unique image processing technique to detect steps [52]. The technique uses video feed collected by the user pointing a camera at their feet as they walk. The algorithm starts by extracting edges with the Canny Edge Detector. The lines in the figure are then eliminated with the Hough Transform. At this point, the features of the image are computed with Fast Features to Track. The center cluster of feature points is determined with K-means, and the x-coordinate of the cluster center is recorded. In this paper, the x-coordinate is in the direction of forward motion of the video. The recorded x-coordinates are smoothed with a Savitzky-Golay low pass filter. After this process is completed, local minima are found, and the number of local minima are reported as the number of steps taken. We cannot implement this algorithm because the video recorded in our

dataset was not recorded in the same way as the proposed algorithm. In addition, this algorithm requires participants to constantly record their own feet while walking, which is impractical in daily life.

Another technique used in step detection is based on conversion of the accelerometer signal to the frequency domain through use of the fast Fourier transform [73]. This conversion technique is followed up by an analysis of the power spectrum. The frequency range of human stepping can be examined in segmented portions of the data. This range is typically near 0.5 to 3.0 Hz. It is common to normalize the maximum power for each window and combine the power spectrums for the x, y, and z axes. The frequency at the maximum power is considered to be the cadence of walking, and the cadence multiplied by the length of time of the segment being examined provide a step count for that segment. Summing up the steps from each non-overlapping segments provides a final step count. Typically, for this type of frequency analysis, the signal should be sampled at a high frequency, for example, 128 Hz [73]. Our dataset was collected at a frequency of 15 Hz, but in order to facilitate this type of frequency analysis, future data can be collected with a significantly higher sample rate. This type of analysis, because of its reliance on periodic signals, is likely only appropriate when applied to regular gaits.

## **3.1.2 Implemented Algorithms**

### **3.1.2.1 Peak Detector**

We implemented the algorithm proposed by Gu et. al. to be representative of peak detection algorithms because it is very recent (2017) and provides enough detail for reimplementing [48]. The algorithm applies preliminary processing to find the magnitude of the signal, identifies peaks, and applies three conditions to each peak in order to determine if it is a step as summarized in Figure 3.5. The algorithm performs signal preprocessing by finding the magnitude of acceleration through use of Equation 2.1. Then, local maxima are detected and the data is segmented based on these peaks. Based on the variance within these samples, the segments are classified as idle, walking, or running. In our implementation of this algorithm, because our dataset only examines walking, we classify the data as either idle or walking only.

In phase two of the algorithm, periodicity, similarity, and continuity of the peaks are scored. The similarity measure compares the magnitude of every other peak. The algorithm examines every

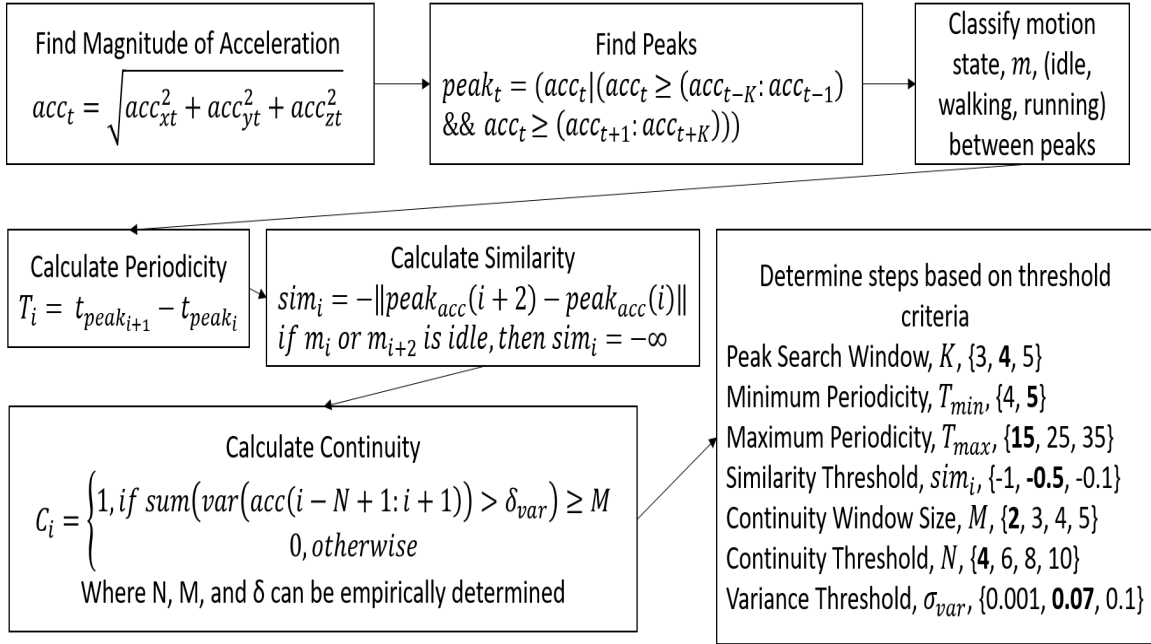


Figure 3.5: Flowchart of the peak detection algorithm. Parameters used in the algorithm are listed, with the values used in the publication in bold.

other peak because these correspond to steps taken with the same foot. For example, if peak 1 is associated with a step taken with the right foot, it is natural to assume that peak 2 would be associated with a step taken by the left foot, and peak 3 would be back to the right foot. This is particularly important when considering the ankle position because peak magnitudes will vary greatly. In order to score periodicity, the time difference between two neighboring peaks is measured and recorded. This value is later compared against a threshold to remove peaks not associated with steps. The continuity of each peak is also scored by counting the number of neighboring peaks that take place during walking (as opposed to idle time segments).

In phase three of the algorithm, thresholds are applied to each peak based on the periodicity, similarity, and continuity scores calculated in phase two. First, the number of neighboring peaks that occur during walking, as calculated as the continuity score, must exceed a threshold for the peak to be considered further. Then, the similarity score, which is the difference in magnitude of alternating peaks, must be less than a threshold. This comparison is based on the concept that if peaks differ significantly, they are more likely to be noise, rather than steps. Finally, the time between consecutive peaks must be greater than a minimum threshold and less than a maximum

threshold in order to be considered steps. These thresholds are based on the frequency of steps occurring in human walking. If a peak passes all three of the specified thresholds, then the peak is considered to be a step.

### 3.1.2.2 Threshold Based

We selected the algorithm developed by Zhao to be our representative sample because the algorithm provided sufficient detail for reimplementing and was representative of multiple zero crossing algorithms [67]. The algorithm's steps are summarized in Figure 3.6. The algorithm first preprocesses the signal by performing a smoothing, essentially averaging every four data points. The algorithm then determines the dynamic zero crossing threshold by finding the maximum and minimum values at 1/2 second intervals. The threshold for the used for detecting zero crossings is the average of the maximum and minimum values found. It is important to note that the values used to determine the threshold for every interval is based on the actual maximum and minimum values calculated for the previous 1/2 second interval. In this way, the maximum and minimum values used to determine the adaptive threshold are delayed for each interval.

After determining the adaptive threshold for each time interval, individual data points are evaluated. This process begins with placing the data into registers, *sample\_new* and *sample\_old*. When a new data sample is considered, the *sample\_new* register is shifted to the *sample\_old* register unconditionally. Then, the new data sample is only placed in to *sample\_new* only if the change in acceleration is greater than a precision value. If the new data sample does not change at least as much as the precision threshold, *sample\_new* is not updated. Then, if there is a negative slope in the acceleration plot (*sample\_new* is less than *sample\_old*) when the acceleration crosses the dynamic threshold, a potential step is detected. This detection process is performed on the x, y, and z axes independently, and whichever axis has the greatest change in acceleration is used to determine if a potential step is present.

The algorithm then adds two conditions to verify whether each potential step is considered to be an actual step. First, a time window is used to rule out vibrations. The time window is based on the assumption that people can run as rapidly as 5 steps per second or walk as slowly as one step per two seconds. Based on this, the time interval between two valid steps must be between 0.2 seconds and 2.0 seconds. Any steps with intervals outside this window are discarded. Subsequently, the algorithm requires that four consecutive valid steps must be found in sequence in order for

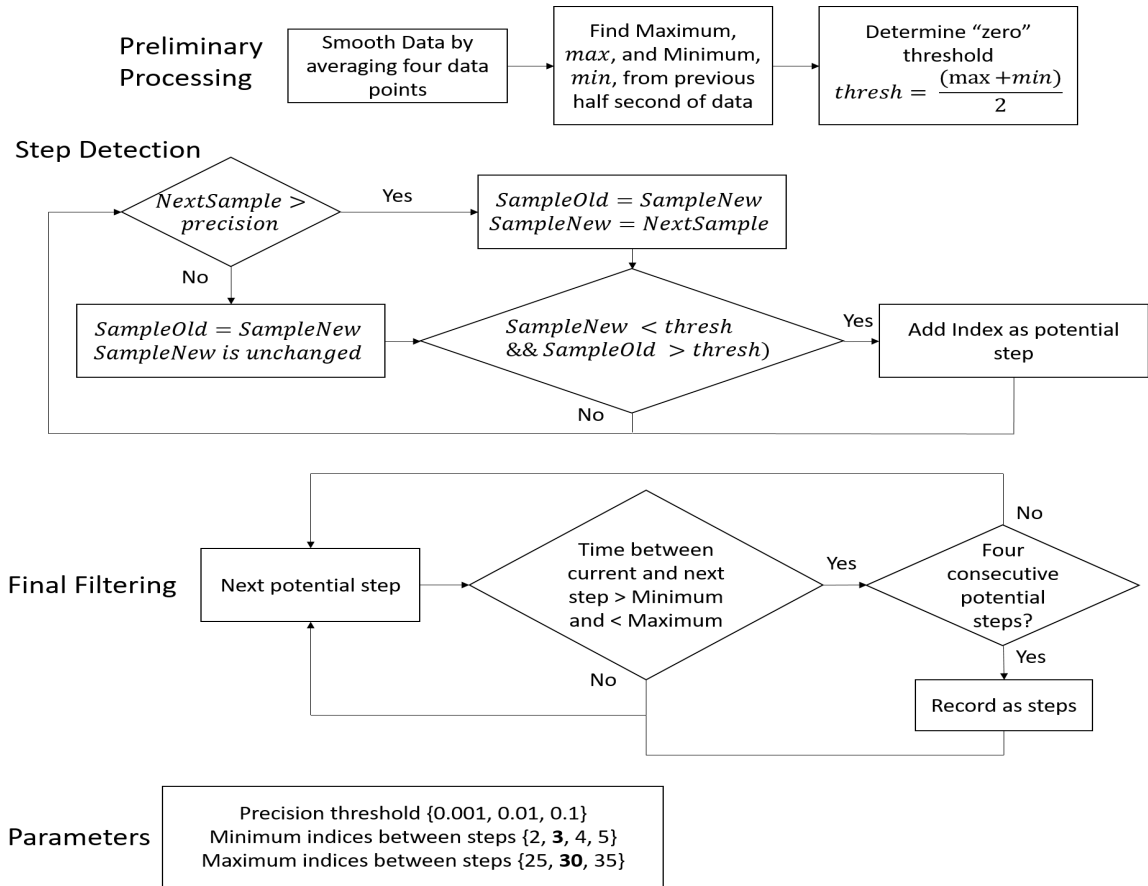


Figure 3.6: Flowchart of the threshold based algorithm. Parameters used in the algorithm are listed, with the values used in the publication in bold.



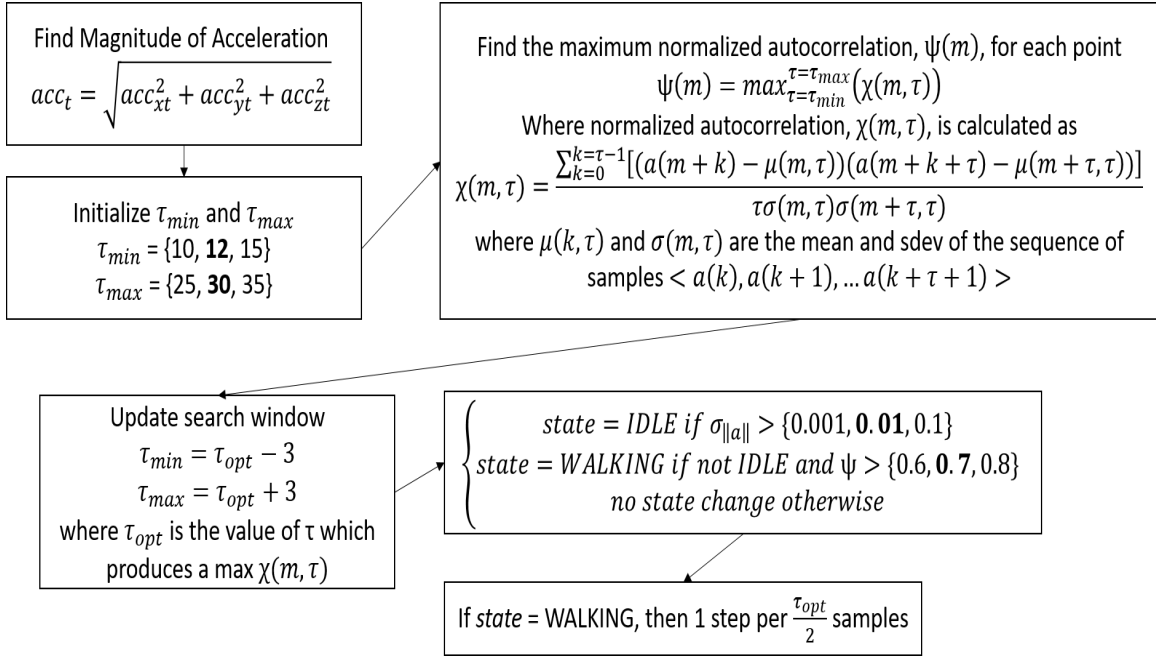


Figure 3.7: Flowchart of the autocorrelation algorithm. Parameters used in the algorithm are listed, with the values used in the publication in bold.

the steps to be counted. If even one invalid step is found, steps are no longer recorded until four consecutive valid steps are found.

### 3.1.2.3 Autocorrelation

Of the two algorithms we found using autocorrelation, we reimplemented the algorithm listed by Rai et. al. [61]. Both papers using this algorithm are from the same research group, so we used the paper providing a greater level of detail regarding algorithm implementation. The algorithm was designed for data collection from a smartphone and is summarized in Figure 3.7. Male participants kept the smartphone in the shirt pocket, front and rear pants pockets, and in hand (while using and not using the phone). Female participants placed the smartphone in a handbag and in hand (while using and not using the phone). These locations should provide similar signals to the hip and wrist sensors used in our data collection process.

The algorithm is based on calculating the autocorrelation at each sample. Autocorrelation is determined based on measures of mean and standard deviation values for intervals of time. This autocorrelation is calculated and normalized for a window size of  $\tau$  for the  $m^{th}$  sample as shown in

Equation 3.1

$$\chi(m, \tau) = \frac{\sum_{k=0}^{k=\tau-1} [(a(m+k) - \mu(m, \tau))(a(m+k+\tau) - \mu(m+\tau, \tau))]}{\tau \sigma(m, \tau) \sigma(m+\tau, \tau)} \quad (3.1)$$

where  $\mu(k, \tau)$  and  $\sigma(k, \tau)$  are the mean and standard deviation of the acceleration samples from  $k$  to  $k + \tau + 1$ . The normalized autocorrelation value approaches 1 as the acceleration in the windows being compared becomes more similar.

The challenge in this process is determining the correct window size, or  $\tau$ , to use for comparing intervals of time. For example, if a participant is walking at a rate of 1 step per second, the appropriate window for autocorrelation evaluation would be 1 second. If, instead, the participant is taking two steps per second, the appropriate window would be 0.5 seconds. Because this value is not known and can change through data collection, the algorithm tests a range of window sizes for each sample being evaluated. For each sample, windows sizes ranging from  $\tau_{min}$  to  $\tau_{max}$  are tested. Whichever window size produces the highest autocorrelation score is used (the window size and autocorrelation value are saved). The values of  $\tau_{min}$  and  $\tau_{max}$  are varied throughout the analysis process. At the start,  $\tau_{min}$  is set to be .8 seconds, or in our case, with a sample rate of 15 Hz, 12 samples and  $\tau_{max}$  is initialized to 2 seconds, or 30 samples. Once a valid step is found by the algorithm, the optimal window size ( $\tau_{opt}$ ) recorded for the sample is then used to adjust the values of  $\tau_{min}$  and  $\tau_{max}$  such that  $\tau_{min} = \tau_{opt} - 3$  and  $\tau_{max} = \tau_{opt} + 3$ .

The final process in determining steps is classifying the state as idle or walking based on the magnitude of acceleration and maximum autocorrelation score found for each data point. If the magnitude of the data point being examined is less than 0.01 gravities, then the state is determined to be idle, so no steps are recorded. Otherwise, if the maximum autocorrelation score found for the data point is greater than 0.7, then the state is determined to be walking. If neither of these conditions are true, then the state does not change. Then, if the state is walking, the optimal window size,  $\tau_{opt}$  for the data point is used to generate a step at every  $\frac{\tau_{opt}}{2}$  samples.

### 3.1.3 Parameter Tuning

Every pedometer algorithm reimplemented in this work relies on multiple parameters which are heuristically determined in order to count steps. Within each prior work, values were assigned and reported for each parameter based on the dataset evaluated. Because sample rate varies between

prior works and this work, some parameters were adjusted accordingly. For example, if a prior algorithm required at least 10 samples between steps with a sample rate of 50Hz, the parameter was adjusted to 3 samples for our sample rate of 15 Hz. Similarly, if the accelerometers used recorded in meters per second squared, and our accelerometer reports values in gravities, parameters were divided by 9.8 in order to convert the values appropriately. This section discusses the default parameters used in each of the algorithms implemented and the process used to tune parameters to the dataset discussed in this work.

The peak detector developed by Gu et. al. [48] used seven parameters which were heuristically determined. The accelerometer used in the work sampled at 50 Hz and provided acceleration in meters per second squared. Our data was collected at 15 Hz with acceleration reported in gravities, so thresholds dealing with the sample rate, number of samples, or acceleration thresholds were converted to match our data. The first parameter identified is used in the initial detection of peaks. The acceleration at each time,  $t$ , is compared to the acceleration values from the  $K$  surrounding data points, and if the acceleration at  $t$  is greater than or equal to the acceleration at all  $t \pm K$  data points, then a peak is considered to be present at time  $t$ . In the paper,  $K$  is determined to be 15 samples, and is converted to 3 samples in our dataset to account for the sample rate differences. In determining the periodicity of the peaks, the original algorithm requires that consecutive peaks be between  $T_{min} = 0.3$  to  $T_{max} = 1$  second apart or they would be removed from consideration as steps. These values are converted to 4 to 15 indices apart at our sensor’s sample rate. The similarity parameter used in the algorithm is set to  $sim_i = -5$ . This value indicates that alternating peak’s acceleration values must be within  $5 m/s^2$  of each other, or for our dataset,  $0.5 gravities$ . In order to evaluate continuity, two parameters are assigned for window size,  $M = 2$ , and number of segments which must be classified as walking  $N = 4$  within the window. In this case, segments are determined based on the time from one peak to another, so no conversion is necessary. The final default parameter in this algorithm is the variance threshold,  $\sigma_{var} = 0.7$  meters per second squared. This value is used to classify segments of data as walking or idle, and is converted to  $\sigma_{var} = 0.07$  gravities for this dataset.

The threshold based algorithm developed by Zhao [67] relies on three parameters. The accelerometer used in the work sampled at 50 Hz and provided acceleration in gravities. Because our data also provides acceleration in gravities, acceleration based thresholds were not adjusted, but sample rate and number of sample thresholds were adjusted to match the sample rate used in our

data collection (15 Hz). The first variable is a precision threshold to determine if acceleration has changed significantly enough to indicate motion. The algorithm does not specify a default precision value, so a default value of 0.01 gravities was selected. The algorithm also uses a minimum time between steps (10 samples) and a maximum time between steps (100 samples), which were converted to 3 and 30 samples, respectively, in our implementation of the algorithm using default parameters.

The autocorrelation algorithm developed by Rai et. al. [61] relies on four parameters. The dataset used an accelerometer which sampled at 50 Hz and reported acceleration in gravities, so we converted parameters relying on frequency to match the 15 Hz frequency used in our work and maintained any parameters relying on acceleration values. The algorithm used an threshold to determine an idle state, which required that the standard deviation of acceleration remain below 0.1 gravities in order to indicate an idle state. This value of 0.1 gravities was used as the default value in our evaluation process. In order to determine the presence of walking, the maximum normalized autocorrelation for a data point must exceed 0.7, according to the algorithm presented. Since this value is unitless, 0.7 was used as our default threshold in this work. Finally, the minimum and maximum autocorrelation periods were specified in the work as 40 and 100 indices. When converted to our sample rate, the default values used in our work were 12 and 30, respectively.

The parameters for each of the three reimplemented algorithms were also trained on the dataset presented in this work for comparison. There were a number of options for parameter tuning. Parameters could be tuned to provide the best performance on any combination of gait type (regular, semi-regular, or unstructured), sensor position (wrist, hip, or ankle), or evaluation metric (RCA or SDA). In this work, parameters were tuned to provide optimal performance on a wrist worn, regular gait sensor according to best RCA. These options were selected because the wrist is the most common location for fitness trackers to be worn, RCA is the most common evaluation metric used in prior works, and regular gait is the most commonly examined gait in prior works. In order to tune parameters, the parameter space surrounding each of the default parameters was searched for each algorithm, according to Tables 3.2, 3.3, and 3.4.

For each of the algorithms, each combination of the parameters was tested in order to find the highest RCA for step detection given the accelerometer signal from the wrist worn sensor during regular gait. In total, 1,620 combinations of parameters were tested for the peak detector, 36 for the threshold based algorithm, and 81 for the autocorrelation algorithm. Parameter space was searched to a similar degree for each parameter, so the peak detector, which relies on seven unique thresholds,

Parameter	Values searched
Peak Search Window, $K$	{3, 4, 5}
Minimum Periodicity, $T_{min}$	{4, 5}
Maximum Periodicity, $T_{max}$	{15, 25, 35}
Similarity Threshold, $sim_i$	{-1, -0.5, -0.1}
Continuity Window Size $M$	{2, 3, 4, 5}
Continuity Threshold, $N$	{4, 6, 8, 10} (if $N < 2M + 1$ )
Variance Threshold $\sigma_{var}$	{0.001, 0.07, 0.1}

Table 3.2: A listing of sample spaces searched for each variable used in the peak detection pedometer algorithm.

Parameter	Samples searched
Precision Threshold	{0.001, 0.01, 0.1}
Minimum Indices Between Steps	{2, 3, 4, 5}
Maximum Indices Between Steps	{25, 30, 35}

Table 3.3: A listing of sample spaces searched for each variable used in the threshold based pedometer algorithm.

Parameter	Samples searched
Idle threshold	{0.001, 0.01, 0.1}
Walking threshold	{0.6, 0.7, 0.8}
Minimum window indices	{10, 12, 15}
Maximum window indices	{25, 30, 35}

Table 3.4: A listing of sample spaces searched for each variable used in the autocorrelation pedometer algorithm.

results in a large number of possible parameter combinations.

## 3.2 Results

This section describes the analysis process for evaluating the accuracies of the three previously developed pedometer algorithms reimplemented in this work. The vast majority of prior works evaluating pedometer algorithm performance use a single metric, which is called Running Count Accuracy, or RCA, in this work. As discussed in Section 2.1.5, when using RCA as an evaluation metric, even if a very high number of false positives and false negatives are present, if they are similar in number, then a high accuracy will be reported. A better metric for evaluation is Step Detection Accuracy, or SDA, which is the F1 score applied to step detection. This takes in to account how many false positives and false negatives are present in the algorithms. In this work, SDA is considered to be a measure of an algorithm’s accuracy, with RCA used to interpret whether the inaccuracies present in each algorithm will result in a long term over- or under- estimation of the number of steps taken.

The process for evaluating the prior pedometer algorithms is particularly challenging because the problem begins with two dimensions (algorithm and gait type), as shown in Figure 3.8a. It is also important to consider sensor position, as this has been shown to impact accuracy. With the inclusion of sensor position, there are 27 combinations possible, as shown in Figure 3.8b. When algorithm parameters and evaluation metric are also considered, the problem grows to include five dimensions, as shown in Figure 3.8c, and all axes have different units. The five axes considered in evaluation are visualized in Figure 3.9. In addition, to these five axes, the analysis could be performed per person or per step. There is also a question regarding whether to treat shifts as steps during the evaluation process or to ignore the presence of shifts. In this work, all metrics are evaluated per person, with RCA, PPV, Sensitivity, and F1 score being calculated for each gait type for each participant, and shifts are counted as steps for evaluation.

Because the dataset used in this work provides accelerometer data for three activities and from three locations on the body, pedometer accuracy can be evaluated based on nine combinations of these variables. In addition, the overall accuracy across activities and across body positions can be calculated by taking the mean of the three categories. Each of the three algorithms implemented and tested on this data set is evaluated for each pedometer location, each activity, across all locations,

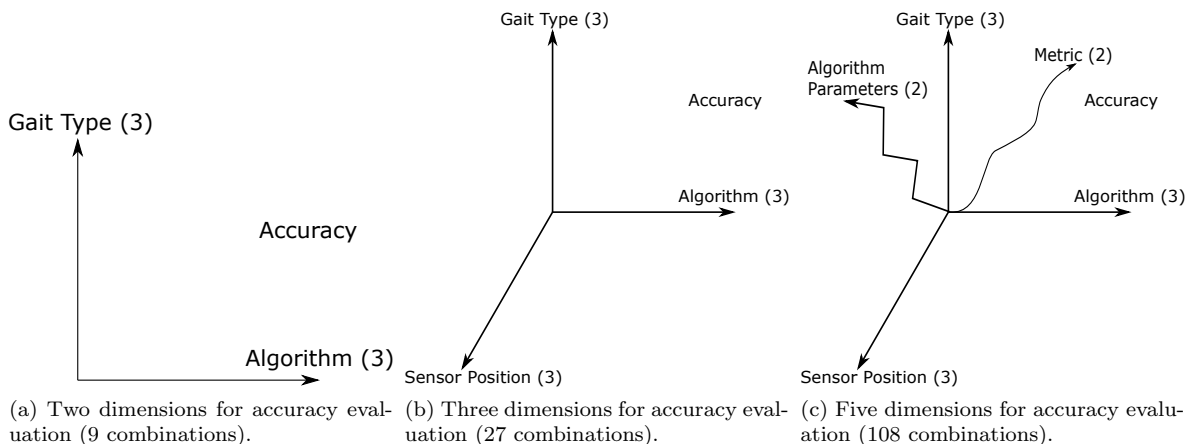


Figure 3.8: Demonstration of the complexity involved in pedometer evaluation, given the dimensionality of the problem.

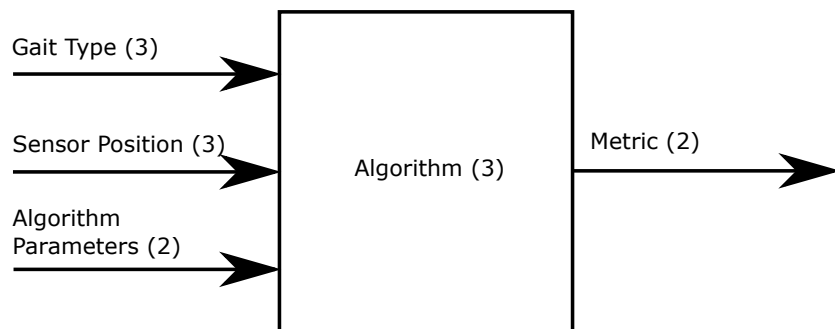


Figure 3.9: The five variables that can be examined and adjusted when evaluating pedometer algorithms on the proposed dataset. The number in parentheses indicates the number of options examined for each variable.

and across all activities. In addition, because RCA and SDA can both be used for evaluation, a total of 24 measures of accuracy can be reported for each algorithm. This process allows each algorithm to be evaluated holistically, to determine its performance across all categories in the dataset.

Each participant is treated as a data point, and all 30 participants' evaluations are averaged for evaluation across each of the five axes. It is worth noting that another variation in evaluation that could be performed is weighting averages based on the number of steps per participant or per category. This work weights each average the same, disregarding the number of steps. Having fixed evaluation to be performed per participant and with equal weighting, there are still three variations in gait type (regular, semi-regular, unstructured), three variations in sensor location (wrist, hip, ankle), two variations in algorithm parameters (default, trained), three variations in algorithm being

examined (peak detector, threshold based, autocorrelation), and two variations in metric being used (RCA and SDA). Providing averages for each of these combinations provides 108 data points for evaluation.

Obviously, a table of 108 results is extremely difficult to interpret and draw conclusions from when examined individually. The following sections will present and examine these results in order to answer 5 questions regarding the data. First, is there a significant difference in performance between the default parameters used in the papers detailing each algorithm and parameters tuned to the dataset? Second, does the metric used for evaluation (RCA or SDA) have a significant impact on the results? Third, how do the performances of each of the three algorithms compare to each other across the full dataset? Fourth, how does the change in gait type affect the accuracy of the pedometer algorithms. Fifth, does the position in which the sensor is placed have a significant impact on step detection accuracy? The analyses across each of these five axes can provide valuable information for the development and evaluation of pedometer algorithms.

### 3.2.1 Evaluation of Parameter Training

In order to evaluate the how tuning parameters effects algorithm accuracy, all combinations of gait type, sensor location, and evaluation metric were considered for each algorithm. The accuracy for each algorithm for each condition are provided in Table 3.5. In this table, accuracies (in the case of RCA) and F1 scores (for SDA) are reported for each of the 108 possible combinations examined. In order to determine whether tuning parameters has an effect on algorithm accuracy, the results using default parameters were compared to results using trained parameters (a value closer to 1 is more accurate). The more accurate result in this comparison is bolded. Due to the number of comparisons which must be made, Table 3.6 provides a simplified result, displaying the number of instances in which the default parameters provided a higher accuracy (14) and in which the trained parameters provided a higher accuracy (40). From this result, it is clear that trained parameters do provide a significant improvement in accuracy.

This result, demonstrating that trained parameters provide higher accuracies for 40 of the 54 comparisons made, is particularly interesting because the parameters were trained for optimal performance on a wrist worn sensor during regular gait using RCA as the evaluation metric. Even tuning to this specific set of conditions, accuracy improves across many categories and combinations. Because the trained parameters performed significantly better than the default parameter set, the



		Default Parameters						Trained Parameters					
		RCA			SDA (F1 score)			RCA			SDA (F1 score)		
Pos	Alg	Reg	Sem	Uns	Reg	Sem	Uns	Reg	Sem	Uns	Reg	Sem	Uns
Wrs	Peak	0.26	0.13	<b>0.06</b>	0.36	0.19	0.09	<b>1.00</b>	<b>1.39</b>	5.88	<b>0.97</b>	<b>0.81</b>	<b>0.30</b>
Wrs	Thr	0.77	<b>0.99</b>	<b>3.35</b>	0.82	0.70	<b>0.29</b>	<b>0.94</b>	1.22	4.52	<b>0.85</b>	<b>0.72</b>	0.28
Wrs	Auto	1.06	1.54	8.11	<b>0.97</b>	0.74	0.21	<b>1.00</b>	<b>0.94</b>	<b>1.36</b>	0.95	<b>0.75</b>	<b>0.60</b>
Hip	Peak	0.62	0.22	0.12	0.75	0.33	0.17	<b>0.98</b>	<b>1.28</b>	<b>1.29</b>	<b>0.98</b>	<b>0.84</b>	<b>0.81</b>
Hip	Thr	0.68	<b>0.91</b>	<b>3.01</b>	0.68	0.63	<b>0.32</b>	<b>1.17</b>	1.4	5.58	<b>0.80</b>	<b>0.70</b>	0.27
Hip	Auto	<b>1.01</b>	1.46	1.57	<b>0.96</b>	<b>0.75</b>	0.57	0.89	<b>0.78</b>	<b>0.64</b>	0.87	0.74	<b>0.60</b>
Ank	Peak	0.58	0.49	0.60	0.73	0.63	0.70	<b>0.77</b>	<b>1.03</b>	<b>1.15</b>	<b>0.87</b>	<b>0.81</b>	<b>0.86</b>
Ank	Thr	0.75	<b>0.89</b>	<b>1.64</b>	0.81	0.71	<b>0.5</b>	<b>0.99</b>	1.20	5.10	<b>0.86</b>	<b>0.74</b>	0.29
Ank	Auto	0.95	1.29	1.30	0.90	0.76	0.71	<b>0.96</b>	<b>0.92</b>	<b>0.99</b>	<b>0.91</b>	<b>0.78</b>	<b>0.74</b>

Table 3.5: Accuracy results across each of the five variables used for evaluating prior algorithms: Parameters (default or trained), evaluation metric (RCA or SDA), gait (regular, semi-regular, or unstructured), sensor position (wrist, hip, or ankle), and algorithm (peak detection, threshold based, or autocorrelation).

Default parameters more accurate	Tuned parameters more accurate
14	40

Table 3.6: A comparison of the number of time default parameters or tuned parameters provided better results.

results provided by the algorithms using trained parameters are used in the evaluations for each of the other 4 axes (metric, algorithm comparison, gait type, and sensor position).

### 3.2.2 Evaluation of Metrics

The two metrics used in evaluation, running count accuracy (RCA) and step detection accuracy (SDA), were compared across algorithm, sensor position, and gait type. SDA is used as a measure of an algorithm’s accuracy, and RCA is used to interpret how the inaccuracies in an algorithm affect the long term step count approximation. For example, if SDA is low, but RCA is high, this indicates that an algorithm is inaccurate, but the number of false positives and false negatives are relatively similar, yielding a good approximation of steps over an extended time period. Because trained parameters were demonstrated to provide better overall performance, the metric evaluation process was performed using only trained parameters.

In order to evaluate each algorithm’s overall performance, SDA and RCA were averaged across all gait types and sensor positions. The results of this evaluation are show in Table 3.7.

Algorithm	Overall RCA	Overall SDA
Peak Detector	1.64	0.81
Threshold	2.46	0.61
Autocorrelation	0.94	0.77

Table 3.7: The effect of evaluation metric on determining algorithm performance.

Sensor Position	Overall RCA	Overall SDA
Wrist	2.03	0.69
Hip	1.56	0.73
Ankle	1.46	0.76

Table 3.8: The effect of evaluation metric on performance according to sensor position

From this table, it can be seen that the autocorrelation algorithm’s RCA is closest to 1, indicating that the number of false positives and false negatives are roughly equal across all other dimensions. The peak detection algorithm has the highest SDA, but, in general, it has significantly more false positives than false negatives, as indicated by an RCA significantly greater than 1. Thus, while the peak detector is the most accurate at detecting steps overall, due to the balancing of false positives and false negatives within the autocorrelation algorithm, the number of steps reported by the autocorrelation algorithm over an extended period of time is the closest to the ground truth in the experiments evaluated in this work. Both evaluation metrics provide valuable insight in to each algorithm’s accuracy.

The relationship between evaluation metric and sensor location is shown in Table 3.8. It can be seen that the overall RCA values are above one for all sensor positions, indicating that, in each case, the number of steps is overestimated. Also, moving from wrist to hip and hip to ankle, the RCA approaches one. This indicates that overall, the three algorithms overestimate steps in each sensor position. The overestimation is least severe when the sensor is worn on the ankle and most severe when the sensor is worn on the wrist. This makes sense because the wrist is likely to have more non-stepping motion throughout the activities performed, including building Legos and opening doors. The change in RCA across sensor position is mirrored by the change in SDA, though the effect is less dramatic. While RCA changes from 2.03 to 1.46 from wrist to ankle, the SDA increases from 0.69 to 0.76 from wrist to ankle. This indicates that the accuracy of each algorithm improves for the ankle, compared to the hip, and for the hip, compared to the wrist. In this case, the trend of overestimation decreases as the accuracy increases across body position.

Gait Type	Overall RCA	Overall SDA
Regular	0.97	0.90
Semi-regular	1.13	0.77
Unstructured	2.95	0.53

Table 3.9: The effect of evaluation metric on performance according to gait type.

The relationship between evaluation metric and gait type is shown in Table 3.9. There is a clear increase in pedometer accuracy as gait becomes more regular, as seen in the increase of SDA from 0.53, for unstructured gait, to 0.77 for semi-regular gait, to 0.90 to regular gait. At the same time, by examining the RCA of 2.95, it can be seen that there is significant overestimation of the number of steps in unstructured gait. This trend is significantly reduced for semi-regular gait, where only a moderate overestimation of steps is present, with an RCA of 1.13. This indicates that not only do the algorithms become more accurate for semi-regular gait, but the number of false positives and false negatives becomes more balanced. Finally, for regular gait, the number of steps reported is actually a slight underestimation, indicated by an RCA of 0.97. Thus, for regular gait, accuracy is significantly greater, and the trend of overestimating steps is eliminated, with slightly more false negatives and false positives.

Through this analysis, it is clear that both evaluation metrics provide valuable insight into step detection performance across algorithm, sensor location, and gait type. When evaluating algorithm performance regardless of sensor location and gait, RCA indicated that an autocorrelation algorithm does the best job of balancing false positives with false negatives, providing a good long term step count, while SDA indicated that a peak detection algorithm would provide the best results. Both RCA and SDA also provide valuable insights toward understanding how the three algorithms perform across body position and gait type. A trend of consistent step overestimation is seen across all body positions, but the algorithms are, in general, most accurate when worn on the ankle and least accurate when worn on the wrist. Also, in unstructured gait, the algorithm accuracies were lowest, and the number of false positives far outweighed the number of false negatives, but in regular gait, algorithm accuracies were highest, and the number of false negatives slightly outweighed the number of false positives.

	Regular	Semi-Regular	Unstructured
Wrist	Peak or Auto (1.00)	Auto (0.94)	Auto (1.36)
Hip	Peak (0.98)	Auto (0.78)	Peak (1.29)
Ankle	Threshold (0.99)	Peak (1.03)	Auto (0.99)

Table 3.10: The most accurate algorithm (peak detection, threshold, or autocorrelation), according to RCA, for each combination of sensor position and gait type. Results use parameters trained on regular RCA for the wrist.

### 3.2.3 Evaluation of Algorithms

When evaluating algorithm performance, it can be useful to determine which algorithm is best for a specific scenario. For example, Fitbit may design a pedometer to be worn as a watch. While some percentage of consumers will fail to comply with recommendations by keeping the sensor in their pocket or wearing it around their ankle, the majority will wear the product as directed, on the wrist. So, while the accuracy for the hip and ankle matter, the algorithm which is most accurate for the wrist is the primary concern in this case. In addition, it may be a primary goal to have the pedometer be as accurate as possible over a long period of time, as would be indicated by RCA, or as credible as possible, in which the ability of the algorithm to correctly identify each step is most important, which is indicated by SDA.

In order to determine which algorithm provides the most accurate step count over time for each combination of sensor position and gait type the best performing algorithm, according to RCA, is presented in Table 3.10. From the table, it can be seen that the autocorrelation algorithm performs best for 5 combinations, the peak detector for 4 combinations, and the threshold based algorithm was best for 1 combination. Based on the gait type being expected and the location on which the sensor is worn, it is possible that any of the algorithms could provide the highest RCA. Thus, deciding which algorithm to use can depend upon which location the device will be worn and which gait type is expected to be most common.

It can be valuable, particularly when considering product credibility, to consider accuracy with regard to detecting individual steps. While a high RCA can be produced with a significant number of false positives (the sensor detects a step when one did not occur) and an equal number of false negatives (the sensor does not detect a step even though one occurred), a high number of false positives and false negatives will significantly decrease SDA. This is important if a user performs a test of credibility by taking a small number of steps and examining the pedometer’s reported

	Regular	Semi-Regular	Unstructured
Wrist	Peak (0.97)	Peak (0.81)	Auto (0.60)
Hip	Peak (0.98)	Peak (0.84)	Peak (0.81)
Ankle	Auto (0.91)	Peak (0.81)	Peak (0.86)

Table 3.11: The most accurate algorithm (peak detection, threshold, or autocorrelation), according to SDA, for each combination of sensor position and gait type. Results use parameters trained on regular RCA for the wrist.

step count. When focusing on credibility, the most accurate algorithm for each combination of gait type and sensor position is shown in Table 3.11. From the table, it can be seen that the peak detector reports the highest SDA for 7 of the 9 combinations of gait type and sensor position. The autocorrelation algorithm performs best for wrist worn unstructured gait and ankle worn regular gait.

From these results, it is clear that selecting the best algorithm is situational. Considering not only gait type and sensor position, but also whether long term accuracy or credibility are more important can affect which algorithm should be selected for optimal performance.

### 3.2.4 Evaluation of Gait Types

In order to determine how gait type affects pedometer accuracy, the RCA values for each algorithm and each sensor location are averaged together. The results from this process are shown in Table 3.12. From these results, it is clear that regular gait is the category which provides the most accurate results, with a RCA of 0.97 averaged across all algorithms and sensor positions. The parameters used in this analysis were tuned to perform best on regular gait, so it is possible that other parameter sets could provide different results. This is an item to be considered in future work. The accuracy for semi-regular gait is 1.13. This indicates that, on average, the algorithms tend to overestimate the number of steps taken during semi-regular gait. This slight overestimation of steps is likely the result of the additional hand motion from non-stepping activities like opening a door, reaching to move or pick up an item, or gesturing. The accuracy for the unstructured gait activity was 2.95. This indicates that, on average, the step count reported by the three pedometer algorithms being analyzed would overestimate the number of steps taken by a factor of 3. This large overestimate is largely due to the wrist activity which occurs while building the Lego toy during the unstructured gait activity. Filtering the non-step related motion during activities like making

	Regular	Semi-Regular	Unstructured
Overall RCA	0.97	1.13	2.95
Overall SDA	0.90	0.77	0.53

Table 3.12: The effect of gait type on accuracy.

	Wrist	Hip	Ankle
Overall RCA	2.03	1.56	1.46
Overall SDA	0.69	0.73	0.76

Table 3.13: The effect of evaluation metric on performance according to sensor position

dinner, folding laundry, or working with tools is an area which can be improved in further pedometer algorithm development.

### 3.2.5 Evaluation of Sensor positions

The position on the body on which the accelerometer or gyroscope is worn can create a significant effect on the signal generated. While each part of the body undergoes periodic motion as steps are taken, the specific motion signals generated can vary significantly. These variations can affect step detection accuracy, as shown in Table 3.13. Because each RCA value is greater than 1, it can be seen that, on average, the algorithms and gaits recorded in this dataset tend to cause an overestimation of steps regardless of where the sensor is worn. However, with a RCA of 2.03 for the wrist worn sensor and an RCA of 1.46 for the ankle worn sensor, it is clear that the step overestimation problem can be significantly decreased if the sensor is worn on the ankle rather than the wrist. The hip worn accuracy is 1.56, which is more similar to the ankle than the wrist. This indicates that the wrist shows the most dramatic increase in step overestimation, which is likely in part due to the motion generated in non-stepping activities. It is far more common for the wrists to be in motion than for the ankle or hip to be in motion when not taking steps.

## 3.3 Conclusions and Future Work

This chapter described the analysis and categorization of pedometer algorithms, the methods used for reimplementing three prior algorithms, and the analysis of the three algorithms with regard to parameter training, sensor position, gait type, and evaluation metric. Previous pedometer

algorithms can be categorized based on the base mechanism used to detect steps. This categorization process revealed that the majority of published step detection algorithms rely on peak or zero crossing (based on an adaptive threshold) detection. Based on this finding, two recent (2010 or more recent) algorithms, a peak detection and a threshold based zero crossing detection algorithm, were reimplemented and tested on the collected dataset. In addition, an autocorrelation based algorithm was implemented for comparison. Algorithm evaluation was performed across 5 axes; algorithm, parameter training, sensor position, gait type, and evaluation metric.

Results were obtained for each axis by collapsing the 108 accuracy results calculated for each combination of the 5 axes examined. Parameter space for each algorithm was searched to find the combination of parameters yielding the highest running count accuracy for the regular gait activity. These trained parameters outperformed the default parameters from the original publications in 40 of 54 combinations of algorithm, gait type, sensor position, and evaluation metric. In order to simplify further evaluation, the trained parameters were used in further analysis. Overall, no single algorithm produced the highest accuracy across all combinations of sensor position and gait type. Thus, ideal algorithm for use in a specific situation is dependent on the combination of gait type and sensor position used. With regard to gait type, as gait type moved from regular to unstructured, accuracy generally decreased regardless of algorithm or sensor position. In addition, accuracy tended to be highest when sensors were worn on the ankle, slightly less accurate when worn on the waist, and significantly less accurate when worn on the wrist. The two evaluation metrics examined were running count accuracy (RCA) and step detection accuracy (SDA), as represented by an F1 score. In general, the same trends across gait type and sensor position were seen, regardless of which metric was used. However, when evaluating differences in algorithm accuracy, the results from the two metrics differed, with a better RCA produced by the autocorrelation algorithm and a better SDA produced by the peak detection algorithm.

A trend of overestimating steps was seen in almost all cases for the three pedometer algorithms implemented in this work. This trend is particularly interesting when compared to the error seen in the Fitbit step detection from the previous chapter. The overall step count recorded by the Fitbit underestimated steps regardless of gait. This contrast between implemented algorithm overestimation and Fitbit underestimation of steps indicates that the algorithm used by Fitbit is either better tuned to filter out non-step activities (at the cost of detecting some steps) or uses a different core algorithm.

Future work includes searching parameter space for each algorithm to find optimal parameters for each combination of sensor position, gait, and evaluation metric. While the results indicate that training parameters to the specific dataset does significantly improve accuracy, it would be interesting to determine if training for a specific combination of sensor position, gait, and metric could identify a set of parameters which works best overall. In addition, future work could examine accuracy for individual participants to identify if specific gait patterns or sensor orientations cause significant changes in accuracy. It would also be interesting to implement a wider variety of pedometer algorithm categories, including gyroscope based algorithms or algorithms based in the frequency domain.



## Chapter 4

# Improving Pedometer Performance Through Detection of Gait Type

Pedometer algorithm performance can vary significantly based on the location of the body that the sensor is worn and type of gait being performed by the wearer. If the sensor position and gait type were known, parameters and algorithms could be selected in order to maximize accuracy. Prior research has focused on identifying the position on which the sensor is worn, but no works have attempted to detect the more interrupted gaits that are frequent in daily life. We have designed experiments to elicit gaits common when exercising, walking around a building, and moving around a room, and we developed a classifier to identify which gait is being performed and select the most accurate algorithm accordingly.

### 4.1 Methods

This section describes the methods used to identify which gait is being performed and to select and switch to the most accurate pedometer algorithm based on the gait detected. Because no prior works have attempted to detect the types of gaits present when walking around a building or around a room, novel techniques were developed, and future of analyses can build on the concepts used in this work. When comparing gaits, it became clear that one feature which consistently differs between regular, semi-regular, and unstructured gait is the regularity of motion.

Regular gait, representing exercise activities like walking around a track or on a treadmill, can be defined by a consistent pace followed for several minutes at a time. In semi-regular gait, representing the motion patterns seen when walking around a building, there are some periods of regular motion (while walking down a hallway, for example) which are broken up when opening doors and can include changes in the rate of motion when walking up or down stairs. In addition, walking around a building includes some periods of walking around a room, which may also elicit a different pace. Thus, while there is some regularity to the motion overall, the pace is spread out over a wider range and is more irregular. In unstructured gait, representative of motions seen when walking around a room (for example, cooking dinner, hunting for keys, or vacuuming), there are periods of no motion, periods of motion without steps, and some periods of walking during which a regular stride is rarely reached. The differences in the regularity of the motion, and therefore acceleration signals, indicates that gait can be identified using the frequency domain.

#### 4.1.1 Fast Fourier Transform

In order to analyze the temporal accelerometer signal in the frequency domain, the Fast Fourier Transform (FFT) was applied to the data gathered in the experiment. An example of the FFT applied to a regular gait is shown in Figure 4.1. The FFT is applied to the combined magnitude of the X, Y, and Z axes. Because the accelerometer data was sampled at 15 Hz, the FFT results have a maximum frequency of 7.5 Hz. By the nature of walking, most information in the frequency domain occurs between 0 and 4 Hz. Because our activities involved walking, the dominant frequency tended to occur near 2 Hz, with harmonics ranging from 1 to 3 Hz. The harmonics can be explained based on participants wearing accelerometers on only one side of the body. Because of this, the magnitude of the signal can alternate with every other step. Any variation in the signal that occurs over 4Hz would require participants to have motions which occur over 4 times per second. Because the fastest pace used in the dataset is a walking pace, no cyclical motion reaches that frequency, so data above 4Hz is not shown in future plots. The FFT was applied to the wrist, hip, and ankle independently.

The FFT for the regular gait activity varies depending on which sensor location is used, but can be characterized by a maximum value occurring at just under 2 Hz, regardless of sensor position, as shown in Figure 4.2. This indicates that the participant's walking rate was at just under two steps per second throughout the regular gait activity. The regularity of the rate of motion leads

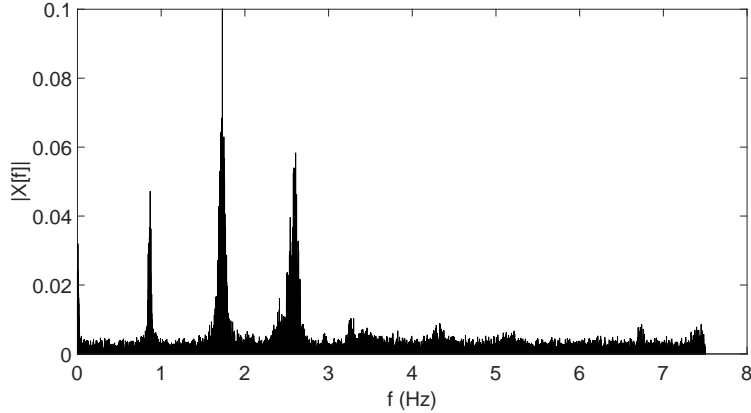


Figure 4.1: An example of an FFT generated by a regular gait activity.

to such a clear spike at the dominant frequency. Across each sensor position, there is at least one clear harmonic frequency.

The FFT resulting from the wrist worn sensor during the regular gait activity is shown in Figure 4.2a. From this figure, it can be seen that the dominant frequency occurs just under 2 Hz, and an additional spike is found at just under 1 Hz. It is possible the dominant frequency is a result of the body’s up and down motion with each step, while the secondary frequency is a result of the forward and backward swing of the arm. The hip completes a motion cycle twice as often because the full torso (and therefore, the arm which is attached) must lift up to allow both the left and right feet to move with each step. When examining the arm motion, however, the swing of the right arm mirrors the motion of the left foot, and visa versa. Thus, since only one arm is instrumented, a full cycle of arm swing motion is only completed every other step. It is also worth noting that because we are using the magnitude of acceleration, the forward and backward motion of the arm each register, causing two accelerations even for one arm cycle, however the magnitude of the acceleration on the back swing is different than the magnitude of acceleration on the front swing. This phenomenon likely contributes to the secondary frequency present at 1Hz for the wrist worn sensor.

When examining the FFT result for the regular gait activity generated by the hip-worn sensor, as shown in Figure 4.2d, it can be seen that the dominant frequency is found at just under 2Hz, with an secondary frequency spike at just under 3Hz. The source of this secondary frequency is not immediately clear. From an analysis similar to the wrist, it can be hypothesized that because the body moves up and down for each step taken, and the magnitude of acceleration is used, there

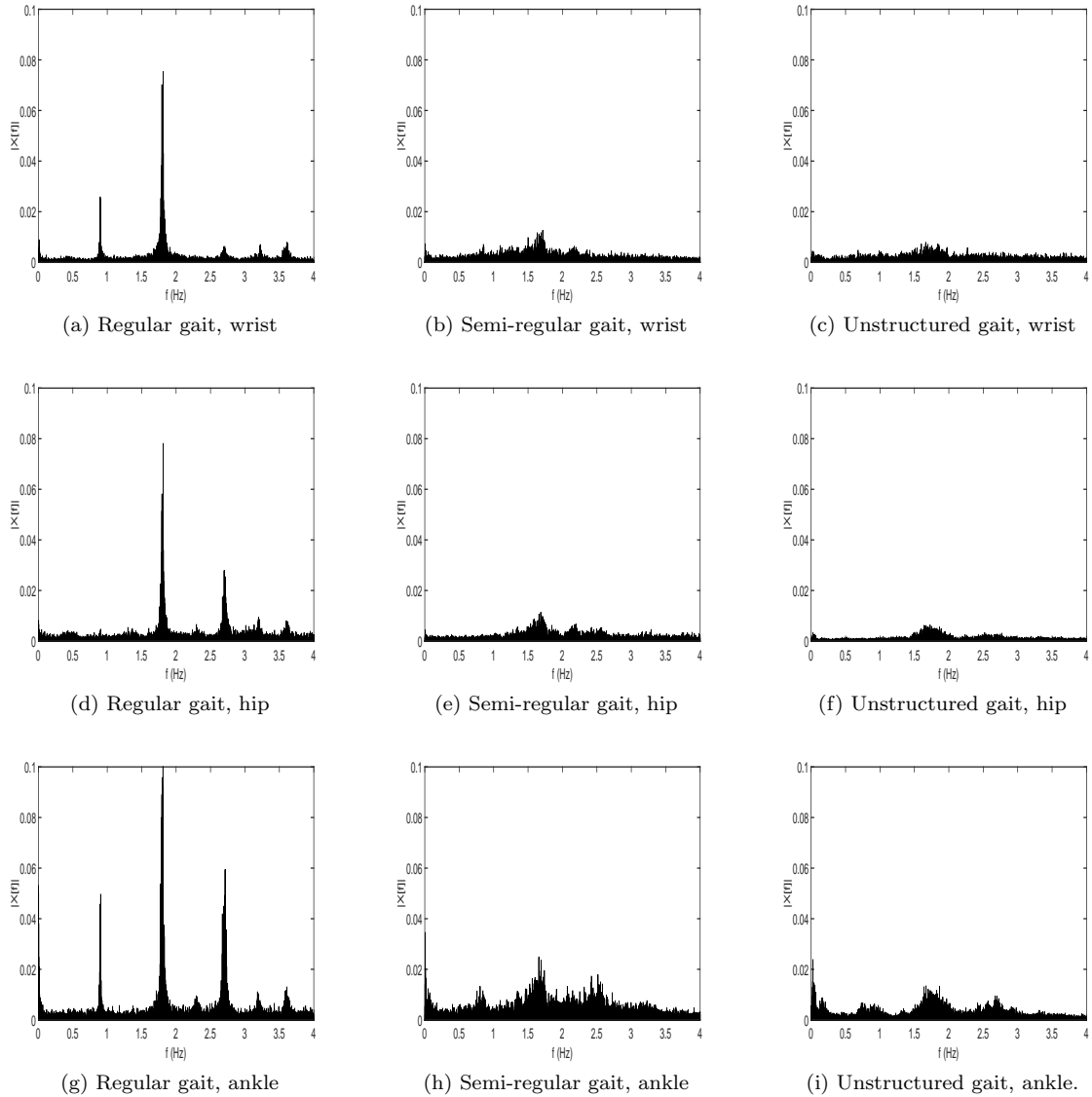


Figure 4.2: FFT results for one participant, for each sensor location (wrist, hip, and ankle) and each gait type (regular, semi-regular, and unstructured). FFT calculated using the magnitude of acceleration (including X, Y, and Z axes).

could be two peaks per step. In addition, the magnitude of the upward and downward accelerations could vary slightly, and the magnitude of acceleration could vary depending on whether a step is being taken with the right foot and the left foot. Thus, secondary frequencies at 1Hz or 4 Hz would be easy to explain, however, the secondary frequency that is actually present in the data at 3Hz is more challenging to understand.

The FFT result for the regular gait activity generated by the ankle-worn sensor, shown in Figure 4.2g, also displays a dominant frequency at 2Hz, and mirrors both the 1Hz frequency spike present in the wrist sensor and the 3Hz frequency spike present in the hip sensor. The ankle-worn sensor only completes one full cycle of motion for every two steps taken. This results in the frequency spike seen at 1Hz. The dominant frequency at 2Hz is a result of the stopping and starting motion which occurs as the foot transitions from rest to motion and motion to rest. The spike at 3Hz is, similar to the hip analysis, challenging to understand. It does seem likely, however, that the phenomenon causing the 3Hz frequency spike in the hip-worn sensor is also the source of the frequency spike in the ankle-worn sensor. One additional feature of note is the spike near 0Hz. This spike is the result of the time period where the foot is stationary while the opposing foot moves.

The FFT was also applied to the wrist-, hip-, and ankle- worn accelerometer signals generated during the semi-regular gait activity, as shown in Figures 4.2b, 4.2e, and 4.2h. When comparing the semi-regular gait FFTs to the regular gait FFTs, it is immediately clear that the dominant frequency, just under 2Hz, is still present, but at significantly reduced magnitude. Where the dominant 2Hz frequency in the regular gait activity has a magnitude between 0.05 to 0.1, the magnitude varies from 0.01 to 0.025 for the semi-regular gait. In addition, the peaks are far more spread out, rather than a clear spike, as in regular gait. This is a result of some regular walking motion as participants walked down hallways, some slower stepping as participants hunted around a room for the hidden object, and some variation in the rate of motion as participants walked up or down stairs. In addition, the 0Hz spike present in the ankle-worn data for the regular gait activity is again present in the semi-regular gait ankle-worn signal (due to each foot remaining stationary as the other moves).

When applied to the unstructured gait activity, the FFT results for each sensor location are shown in Figures 4.2c, 4.2f, and 4.2i. The FFT output for the unstructured gait activity mirrors that of the semi-regular gait activity, but with even further reduced magnitudes for each of the peaks. In fact, it is difficult to clearly pinpoint a dominant frequency in the wrist-worn sensor, and a frequency just under 2Hz is just noticeable in the hip-worn sensor. For the ankle-worn sensor, it

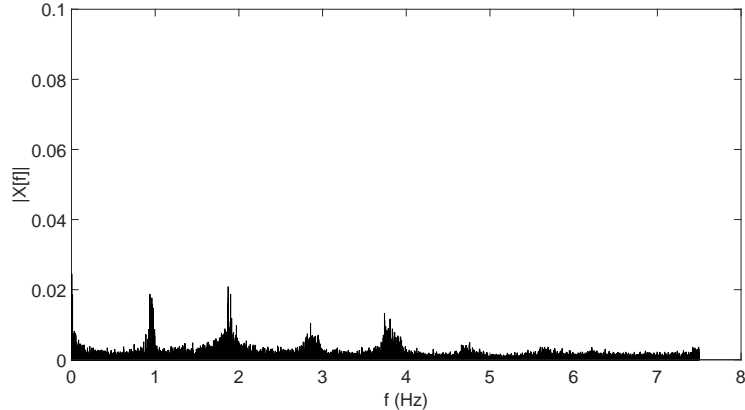


Figure 4.3: An example of an FFT generated by a regular gait activity, wrist-worn sensor. This FFT is difficult to distinguish from a semi-regular gait FFT.

is unclear exactly where the dominant frequency is because there are roughly four peaks near 0Hz, 1Hz, 1.8Hz and 2.5 Hz.

The trends seen in Figure 4.2 generally hold across participants, but not every participant’s FFT results are as clearly separable. The FFT shown in Figure 4.3 is an example of a regular gait activity from a wrist worn sensor which was misclassified as semi-regular gait by our classifier. It is challenging to differentiate the FFT shown with, for example, the ankle-worn semi-regular gait FFT from another participant shown in 4.2h. Because this analysis is focused on how detecting gait could improve pedometer accuracy, rather than considering how sensor position detection could improve accuracy, the sensor location ground truth is not used to aid the classification process. However, accurate sensor location detection could potentially be used to improve gait detection.

### 4.1.2 Gait vs Stride

One goal of this work is to improve pedometer accuracy and believability based on the detection of which gait type is being performed. In order to provide real-time feedback to users, detection of gait should occur at a faster rate than an activity-level detection. In our experiment, each activity was normalized to take roughly 10 minutes to perform. If a pedometer algorithm can only identify which activity was performed based on the analysis of 10 minutes of data, the believability of the algorithm will not be significantly improved. Because of this, we sought to classify much smaller windows (3 to 30 seconds) of motion. We call this analysis stride analysis

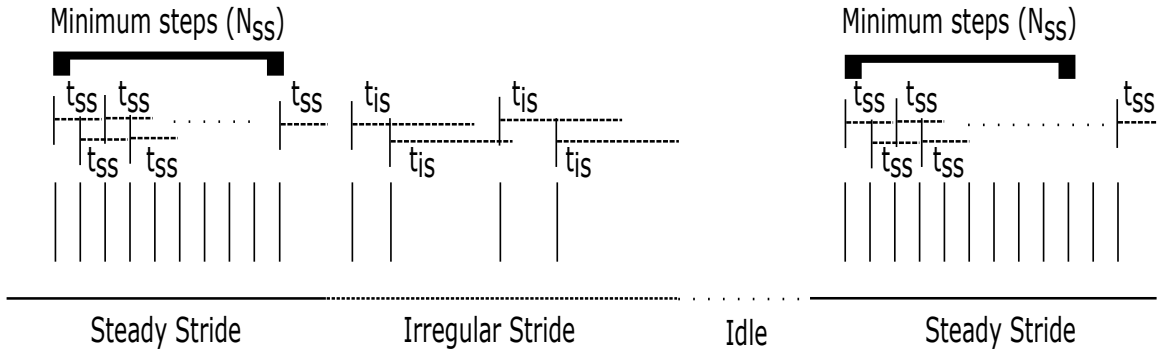


Figure 4.4: An example of the process of creating a ground truth for stride detection (steady stride, irregular stride, idle). The process is based on the ground truth step indices. There are three thresholds for detection: time between steady stride steps ( $T_{ss}$ ), number of consecutive steps needed for steady stride ( $N_{ss}$ ), and time to transition between irregular steps and idle ( $T_{is}$ ). Each of these thresholds were determined heuristically.

because the strides detected differ from the gaits detected.

In order to create a stride detection algorithm, a ground truthing process was developed as shown in Figure 4.4. The ground truth step times, as identified in Section 2.1.4, were used in order to determine whether the participant was in a steady stride, irregular stride, or idle state. Two thresholds were required for a step to be considered a steady stride step: the time between steps was required to be less than  $T_{ss}$ , and  $N_{ss}$  consecutive steps were required to meet the time threshold. Based on a heuristic analysis,  $T_{ss}$  was set as 1 second, and  $N_{ss}$  was set to 10 steps. Any ground truth steps which did not meet the requirements were labeled as irregular stride. A third threshold,  $T_{is}$  is used to define the boundary for irregular stride. This value was determined heuristically and set to 3 seconds. Therefore, any index within 3 seconds of an irregular stride step was labeled as irregular stride. Any remaining unlabeled indices were identified as idle.

In order to ground truth an extended window of time, the majority vote of indices within that window were used, as shown in Figure 4.5. In the figure, a window size of 5 seconds is approximated, but a variety of window sizes, including 3, 5, 8, 10, 15, 20, and 30 seconds were tested in the evaluation process. In the example, 7 of the 8 windows have a clear ground truth value because the majority of the window clearly can be defined as one particular stride type. One challenge with non-overlapping windows, however, is that in some cases, a slight variation in step indices can change a window's ground truth if the window spans the border of two different stride types. This type of borderline classification is demonstrated in the second to last window of the figure. In this window, roughly

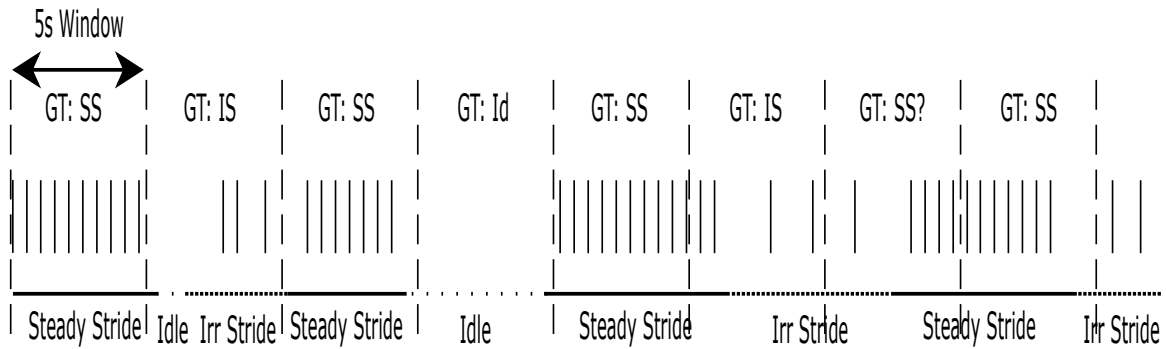


Figure 4.5: A demonstration of the window-based ground truth determination process. In the example, a 5 second window is approximated, and the window’s ground truth value is determined by a majority vote of individual index ground truth values. In the second to last window shown, the window’s ground truth value is very close to either irregular stride or steady stride, demonstrating how, in some cases, small changes in one or two step indices can change the ground truth value for a window.

half of the indices are defined as irregular stride, and the other half are defined as steady stride.

In order to classify the accelerometer data algorithmically, FFTs were calculated for each non-overlapping window. Several examples of FFTs calculated from a semi-regular gait activity are shown in Figure 4.6. In this figure, it can be seen that the semi-regular gait activity is not defined by irregular stride, but rather that multiple strides are combined within the activity. Figures 4.6a - 4.6e demonstrate periods of time in which the participant used an irregular stride while hunting for an object within a room. Figures 4.6f - 4.6g demonstrate periods of time where the participant used a steady stride after finding the object and walking to exit the room. Figure 4.6h demonstrates time spent by the participant opening the door and preparing to exit the room. Figure 4.6i demonstrate steady stride as the participant walks down a hallway toward the stairs. Figures 4.6j - 4.6k include time where the participant is walking up a stair well. In figure 4.6l, the participant resumes walking along a hallway. As demonstrated in the figure, the FFTs for each window of time differ depending on the stride type being performed.

When examining the regular gait activity, we found that, with the exception of the start and end of the activity, where the sensor is active but the participant has not begun walking (or has finished their walk), the stride could be described as a steady stride. There is a clear relationship between steady stride and regular gait. As described previously, the semi-regular gait activity is composed of steady stride, irregular stride, and idle at various points throughout the activity. Similarly, the unstructured gait activity is characterized by significant periods of inactivity (from



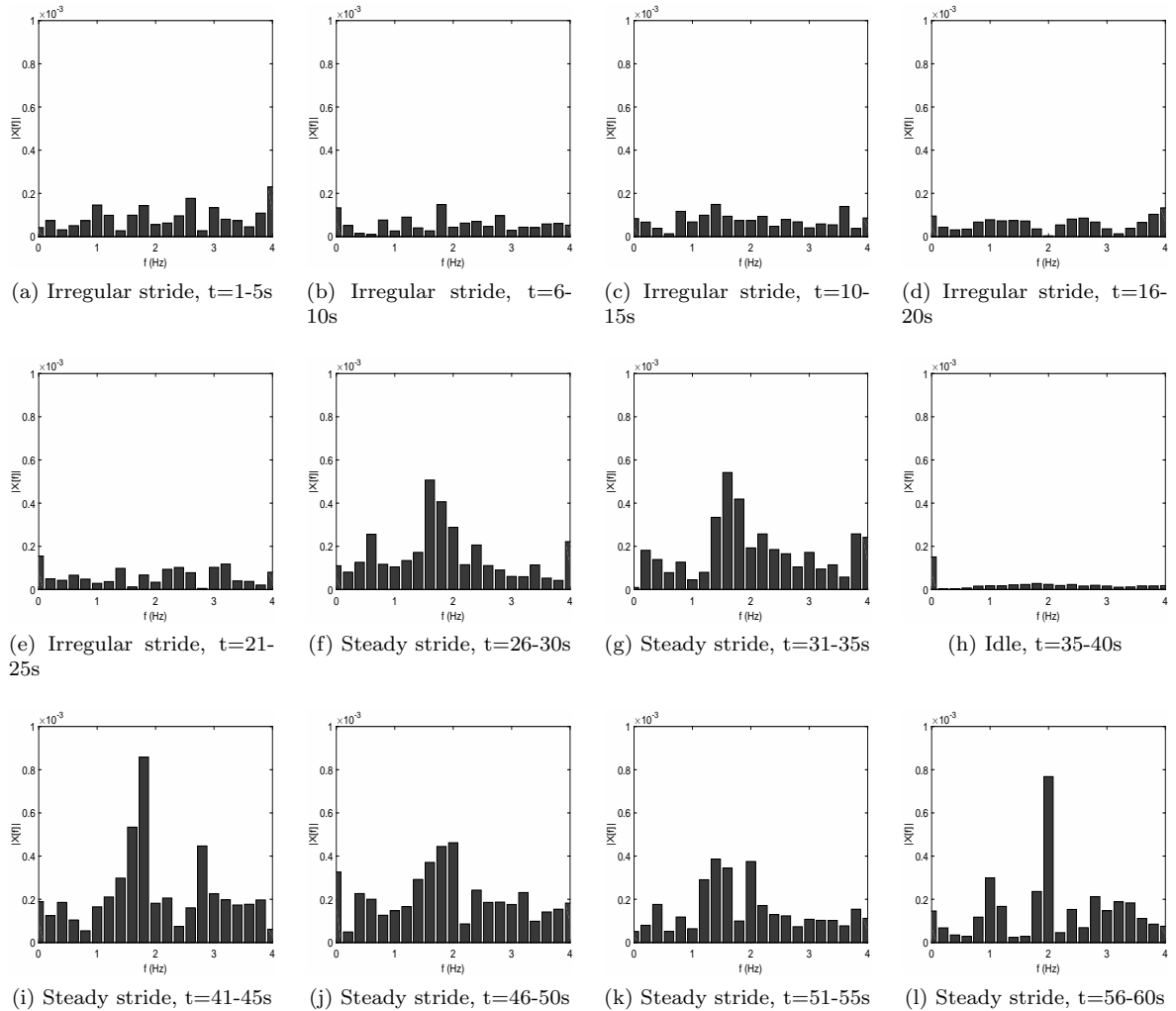


Figure 4.6: Example of 1 minute of motion in a semi-regular gait activity. Motion is broken in to 5 seconds windows and each window of activity is converted to the frequency domain. For the first 25 seconds, the participant is hunting for an object in a room. From 26 - 35 seconds, the participant walks to the door to exit the room. From 35-40 seconds, the participant opens the door. From 41-45 seconds, the participant begins walking down the hallway. From 45-55 seconds, the participant enters the stairwell and climbs stairs. From 55-60 seconds, the participant resumes walking down the hallway.

the perspective of a step counter) as the participant builds the Lego toy, punctuated by brief periods of walking which border between steady stride and irregular stride, as the participant walks to acquire more Lego pieces. The key finding here is that if the window size of gait classification is reduced to smaller time frames, the strides examined are identified separately from the gaits examined, and the gaits being evaluated are potentially composed of multiple segments of each stride type.

### 4.1.3 Classification

In order to identify the activity and stride type through analysis of the accelerometer data, a Naive Bayes classifier was used. The Naive Bayes classifier is a probabilistic classifier which assigns the most probable class  $c_i \in C$ , given feature values  $f_1, f_2, \dots, f_N$ . The classifier uses a naive assumption of independence in features, so the classification problem can be written as shown in Equation 4.1.

$$c_i = \underset{C}{\operatorname{argmin}} P(c_i) \prod_j P(f_j | c_i) \quad (4.1)$$

Our problems each consist of three classes. When determining activity, the classes are regular gait, semi-regular gait, and unstructured gait. When determining stride, the classes are steady stride, irregular stride, and idle. In each case, the three classes are identified as  $c_0, c_1, \text{ and } c_2$ , with  $P(c_i) = 0.33$  for each class.

The features for each class were based on a peak detection algorithm applied to the FFT outputs for each set of data analyzed. The peak detector used for this purpose was a modified version of the peak detector used in the reimplemented peak detection based step detector described in Section 3.1.2.1. In order to reduce the number of false positive peaks, the peak value was required to be at least two standard deviations above the mean of the data. When a value meeting these criteria was found, a range of .2 Hz was searched to find the maximum value. Once a maximum value was found, the value and frequency for the peak were recorded, and no additional peaks could be found within  $\pm 0.1 \text{ Hz}$  of the recorded peak. Through this peak detector, three features were used in order to classify activity type and stride type: the value of the maximum peak magnitude, the number of peaks identified, and the width of the maximum peak. The width of the maximum peak was determined by the minimum and maximum frequencies within  $\pm 0.1 \text{ Hz}$  which exceeded two standard deviations above the mean in magnitude.

In order to apply the Naive Bayes classifier to the data, the means and standard deviations of each of the three features for each of the three classes being examined were found. Then, Equation 4.2, was applied to each segment of data being analyzed in order to predict which class was present.

$$P(f_j|c_i) = \frac{1}{\sqrt{2\pi\sigma_{i,j}^2}} \exp\left(-\frac{(f_j - \mu_{i,j})^2}{2\sigma_{i,j}^2}\right) \quad (4.2)$$

Where  $\mu_{i,j}$  and  $\sigma_{i,j}^2$  are the mean and variance for feature  $j$  for class  $i$ . Based on this classification process, the activity and stride type were predicted for each combination of sensor location, gait type, and window size.

#### 4.1.4 Varying Pedometer Algorithm Based on Detected Gait

The goal of detecting gaits is to improve the accuracy of step detection by selecting the algorithm based on the gait detected. Based on our the analysis of peak detection, threshold crossing, and autocorrelation based algorithms, summarized in Section 3.2.3, we found that identifying the most accurate pedometer algorithm depends on the sensor position and gait type. If body position and gait type are known, then the most accurate algorithm can be selected, and the overall step detection accuracy increased. In order to measure the potential for increasing accuracy based on gait detection, it was assumed that sensor position was known. The knowledge of sensor position, combined with the detected gait type were used to select the algorithm that should be used for step detection. It was also assumed that the manufacturer would determine whether RCA or SDA would be the most important metric for their device. Thus, if SDA was identified as the most important metric, the sensor was worn on the wrist, and semi-regular gait was detected, based on Table 3.11, the peak detection algorithm would be selected. In the same situation, if the most accurate RCA was desired, then according to Table 3.10, the autocorrelation algorithm would be selected. Based on these selections, the accuracy of this varying algorithm pedometer was determined.

## 4.2 Results

This section describes the results obtained for detecting regular, semi-regular, and irregular gait using the Bayes Classifier described in this chapter. The results of detecting steady stride, irregular stride, and idle states are also found for window sizes ranging from 3 to 30 seconds. In

Ground Truth \ Detected	Regular	Semi-regular	Unstructured
Regular	<b>0.93</b>	0.067	0
Semi-regular	0.022	<b>0.81</b>	0.17
Unstructured	0	0.21	<b>0.79</b>

Table 4.1: Confusion Matrix for Gait Detection.

addition, the accuracy of a varying-algorithm based pedometer was identified and reported. Details of performance across gait type and sensor position are also provided. Because stride was not considered when reimplementing prior pedometer algorithms in Chapter 3, no evaluation of a varying algorithm pedometer based on stride detection is reported.

### 4.2.1 Gait Detection

The overall accuracy at detecting regular, semi-regular, or unstructured gaits is 84%. In order to determine where gait detection is challenging, a confusion matrix for gait detection was created, as shown in Table 4.1. It can be seen that regular gait was correctly identified the most often with an accuracy of 93%. The matrix indicates that when a participant was performing a regular gait activity, the classifier never misclassified the activity as unstructured gait, and misclassified regular gait as semi-regular gait 7% of the time. Semi-regular gait was correctly identified with 81% accuracy. When an error was made, semi-regular gait was misidentified as regular gait 2.2% of the time and as unstructured gait 17% of the time. This indicates that there is significant confusion between semi-regular and unstructured gaits. Unstructured gait was correctly identified 79% of the time. When misclassified, it was only misclassified as semi-regular gait, which occurred 21% of the time. This further reinforces the idea that it is more challenging to differentiate semi-regular and unstructured gaits than it is to identify regular gait.

In order to further understand challenges in detecting gaits, the accuracy of gait detection was also broken down according to sensor position, as shown in Table 4.2. From this table, it can be seen that the wrist and hip signals were classified correctly 88% and 89% of the time respectively, while the ankle signals was classified correctly only 77% of the time. Because the classifier was built without information about sensor location, this indicates that the features present in the ankle signal may differ from the features in the wrist ankle across gait types. With knowledge of sensor location, the algorithm for gait detection may be able to improve classification based on accounting

	Wrist	Hip	Ankle
Accuracy	0.88	0.89	0.77

Table 4.2: Gait detection accuracy with respect to sensor location.

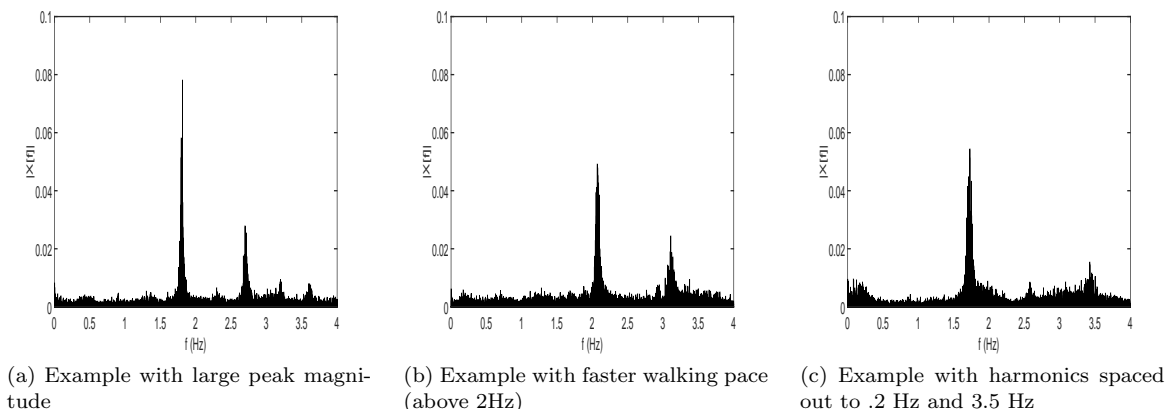
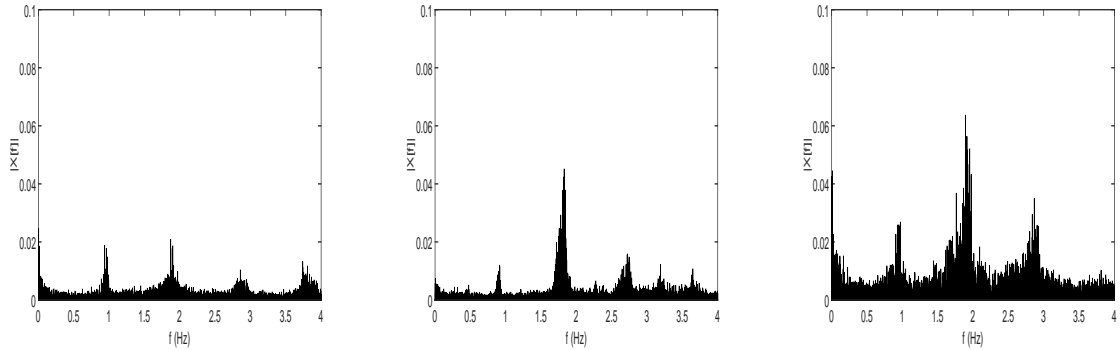


Figure 4.7: Examples of variations within regular-gait, wrist worn FFTs. The variations shown do not cause misclassification.

for these differences.

Within regular gait, there are a number of variations within the resulting FFTs which were correctly classified, as shown in Figure 4.7. The sub-figures demonstrate FFTs resulting from a wrist-worn sensor in regular gait. Figure 4.7a demonstrates a dominant frequency with a peak magnitude of approximately 0.08. The dominant frequency occurs at 1.8 Hz, and a harmonic frequency can be seen at 2.7 Hz. Figure 4.7b demonstrates variation in the magnitude and frequency value of the dominant frequency. The magnitude is only 0.05 Hz in this figure, and the frequency is approximately 2.2 Hz. The resonant frequency is located at approximately 3.2 Hz. The participant with this FFT walked at a faster rate than the participant in 4.7a and exhibited a greater variation in walking rate, resulting in a lower maximum peak magnitude. Figure 4.7c demonstrates how the harmonics in the regular gait FFT can be spaced out differently by participant. This participant had a dominant frequency at 1.75 Hz with weaker harmonics at 0.2 Hz and 3.5 Hz. These variations in FFT did not cause misclassification of the data.

Examples of challenging variations in FFTs which did result in misclassification are demonstrated in Figure 4.8. Figure 4.8a shows an FFT resulting from a regular gait activity using a wrist-worn sensor. The participant in this example demonstrated significantly reduced arm swing



(a) Example of a regular gait activity from a wrist-worn sensor which was misclassified as semi-regular gait. The participant demonstrated very little arm swing while walking.

(b) Example of a regular gait activity from a hip-worn sensor which was misclassified as semi-regular gait.

(c) Example of a semi-regular gait activity seen from an ankle-worn sensor. The FFT was misclassified as regular gait. This participant was the fastest to locate each hidden object and spent a greater proportion of their time with a steady stride.

Figure 4.8: Examples of variations across gait type and body position which resulted in misclassification of activities.

when walking, as compared to other participants. In Figure 4.8b, the FFT was generated from a wrist-worn accelerometer during a regular gait activity. While the characteristic dominant frequency peak is demonstrated, the number of harmonic peaks and the width of the dominant frequency peak caused a misclassification. These feature changes are likely the result of the participant having slight variations in their walking speed throughout the activity. Based on video review, these variations appear to have been the result of avoiding other pedestrians or uncertainty in the path to be followed. In Figure 4.8c, an ankle-worn sensor during semi-regular gait is incorrectly classified as a regular gait activity. This participant was the fastest to find each of the hidden objects, and so they spent a much larger proportion of their time walking with a steady stride up and down hallways.

## 4.2.2 Stride Detection

One critical issue when evaluating stride detection accuracy relates to determining which window size is ideal. A smaller window contains less data to make a classification, but it is also able to adjust more quickly to changes in stride. In order to evaluate where these competing concepts result in an ideal window size, window sizes of 3, 5, 8, 10, 15, 20, and 30 seconds were evaluated. The stride detection accuracies of each window size are shown in Figure 4.9. From this figure, it can be seen that the highest accuracy, 82%, is found with a window size of 5 seconds. This indicates

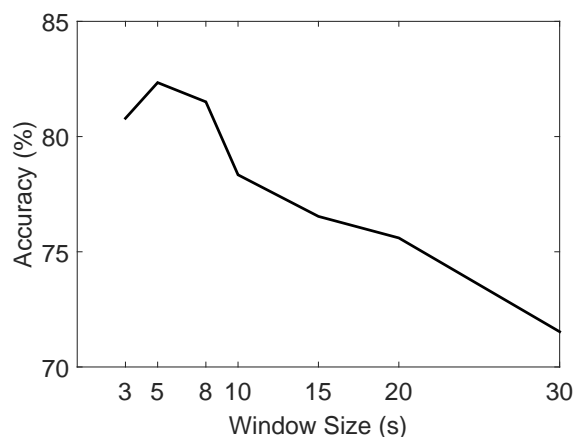


Figure 4.9: Stride detection accuracy as a function of window size.

	Wrist	Hip	Ankle
Accuracy	0.79	0.83	0.85

Table 4.3: Stride detection accuracy using a 5 second window with respect to sensor location.

that a 3 second window does not contain enough data to accurately identify the stride type, but windows of 8 or more seconds are not able to adapt as quickly as a 5 second window to changes in stride. Because of these findings, a window size of 5 seconds was used for further stride detection accuracy evaluations.

Stride detection accuracy was evaluated across sensor location using a 5 second window, as shown in Table 4.3. From this table, it can be seen that the accuracy of stride detection across body position is roughly even. There is some trend of increasing accuracy from wrist to hip and hip to ankle, with detection accuracy increasing from 79% for the wrist to 83% for the hip and 85% for the ankle. The slightly lower accuracy in the wrist data is to be expected due to the increased noise generated by hand motions while building the Lego toy during the unstructured gait activity. In general, however, stride type can be consistently measured across each sensor position.

The confusion matrix for stride detection using a 5 second window is shown in Table 4.4. From this table, it can be seen that steady stride is identified with an accuracy of 95% and idle stride is also identified with a high accuracy, 93%. It is extremely rare for steady stride to be misclassified as idle (1%) or idle to be misclassified as steady stride (2%). The windows which are misclassified in this way appear to always result from occurrences in which nearly half of the window’s ground truth indices are regular gait, and the other half are idle. This type of edge case causes the resulting

Ground Truth \ Detected	Steady stride	Irregular stride	Idle
Steady stride	<b>0.95</b>	0.042	0.01
Irregular stride	0.39	<b>0.21</b>	0.40
Idle	0.02	0.04	<b>0.93</b>

Table 4.4: Confusion matrix for stride detection.

	Ground Truth Time (min)
Steady stride	510
Irregular stride	173
Idle	388

Table 4.5: The ground truth time for each stride type, in minutes.

FFT to border between the two stride types. The most frequent misclassification of stride occurs with irregular stride, which has an accuracy of only 21%. It is worth noting that irregular stride is misclassified as steady stride and idle at roughly the same rate.

The sharp contrast in stride detection accuracy using a 5 second window led to an examination of stride detection accuracy for each stride type across each window size. The detection accuracy for each stride type was examined across window size, as shown in Figure 4.10. It can be seen that the accuracy of steady stride detection increases from window sizes of 3 seconds to 8 seconds, then drops significantly as window size increases from 8 to 10 seconds and remains relatively constant for window sizes from 10 to 30 seconds. Irregular stride detection is very low for 3 to 8 second windows, increases sharply from 8 to 10 second windows, and gradually increases from 10 to 30 second windows. Idle detection accuracy gradually decreases as window size increases. If each stride type were equally represented, a 15 second window size would yield the greatest overall accuracy, however, the stride types are not equally represented. As shown in Table 4.5, steady stride is the most heavily represented stride, with 510 minutes of ground truth indices labeled as steady stride. Idle is represented with 388 minutes of ground truth indices, and irregular stride is represented with 173 minutes of ground truth. The heavy emphasis on steady stride and idle result in the 5 second window having the highest overall accuracy.

The trends seen in stride detection accuracy for the 5 second window are reflected in accuracy across gait type, as seen in Table 4.6. From this table, it can be seen that the regular gait activity was classified with an accuracy of 98%. Because steady stride is extremely common in regular gait,



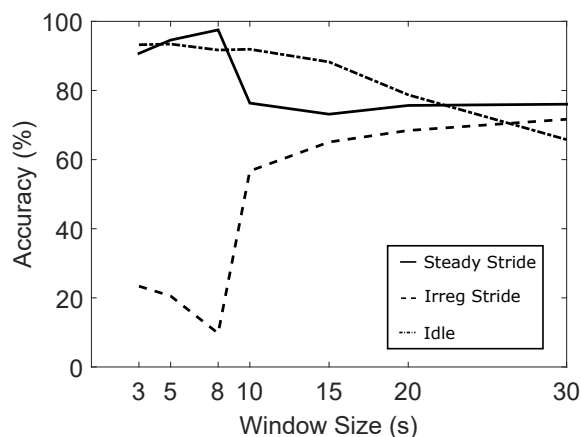


Figure 4.10: Stride detection accuracy as a function of window size, with each stride type examined individually.

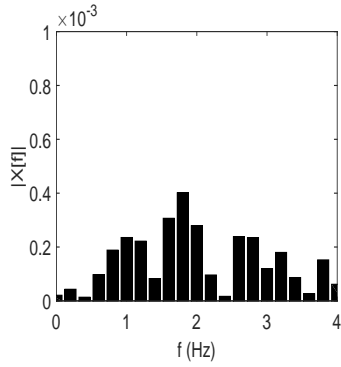
	Regular	Semi-regular	Unstructured
Accuracy	0.98	0.70	0.79

Table 4.6: Stride detection accuracy using a 5 second window, categorized by gait type.

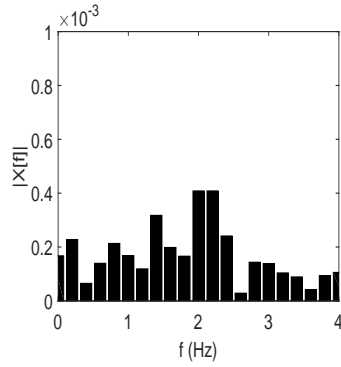
it is to be expected that the strides present in regular gait are so accurately identified. The stride types present in the semi-regular gait activity were correctly classified only 70% of the time. This reflects a higher prevalence of irregular stride within this gait type, although there is still a heavy presence of steady stride and idle. The strides present in unstructured gait are correctly identified 79% of the time, indicating that irregular stride is also present in unstructured gait activities, but at a less common rate than in semi-regular gait activities. Overall, the accuracy of stride detection in gait types reflects the expected presence of each stride type within each activity.

In order to determine when the classifier is failing to correctly identify irregular stride instances, some examples of correct and incorrect classifications of irregular stride were found and are shown in Figure 4.11. Three correctly classified instances of irregular stride are shown in Figures ?? - ?. These examples all have some activity in the frequency domain, generally centered near 2 Hz. Figure 4.11c has the most distinctive maximum peak, and this FFT falls near the boundary of steady stride, but is correctly classified as irregular stride.

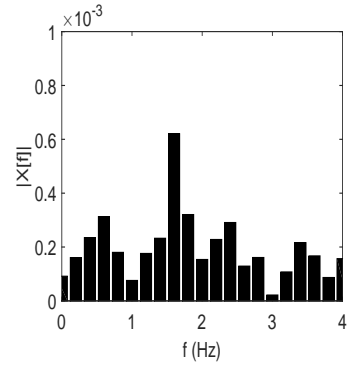
To contrast, three examples of irregular stride FFTs which were misclassified as steady stride are shown in Figures 4.11d - 4.11f. All three of the figures share characteristics similar to the steady stride FFTs. There is a clear dominant frequency near 2 Hz in Figures 4.11d and 4.11e.



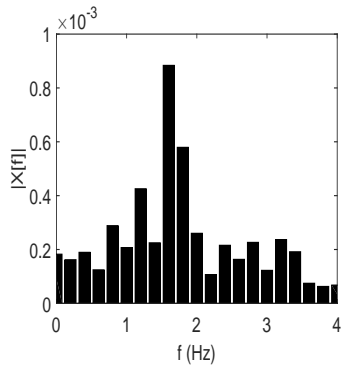
(a) Correctly classified irregular stride.



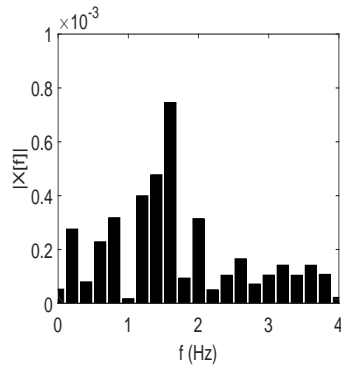
(b) Correctly classified irregular stride.



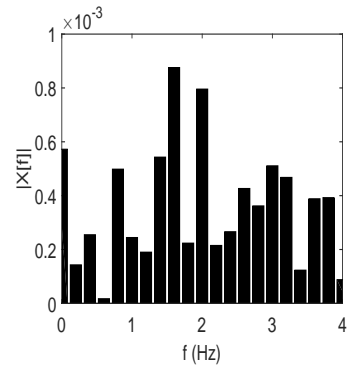
(c) Correctly classified irregular stride.



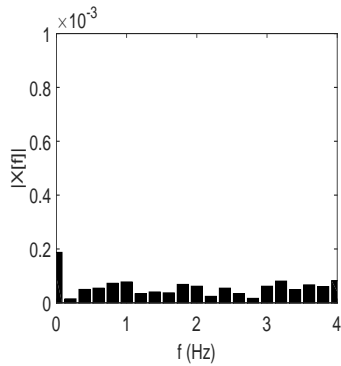
(d) Irregular stride misclassified as steady stride.



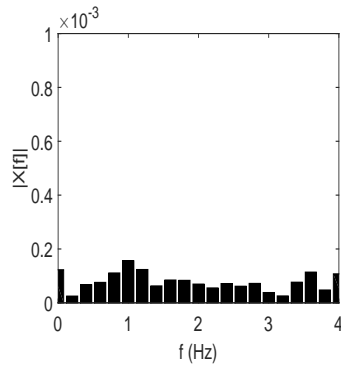
(e) Irregular stride misclassified as steady stride.



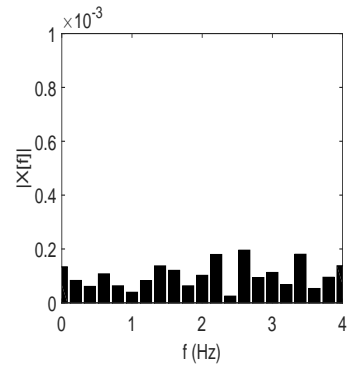
(f) Irregular stride misclassified as steady stride.



(g) Irregular stride misclassified as idle.



(h) Irregular stride misclassified as idle.



(i) Irregular stride misclassified as idle.

Figure 4.11: Irregular stride FFTs were examined to determine why the accuracy rate was low when using a 5 second window for analysis. Examples of correctly classified and misclassified FFTs are shown.

While there is other frequency noise present, due to the shorter 5 second window of time used to gather this data, there are more opportunities for rarely occurring signals to cause noise in the FFT result. This increase in noise is particularly clear in the misclassified signal shown in Figure 4.11f. The FFT shown has two strong frequencies near 2 Hz, and is characterized by a significant number of activity throughout the 4 Hz window. Overall, the FFTs shown in the three plots are difficult to separate from steady stride FFTs and indicate that it may be a challenging problem to separate the two classes given a 5 second window size.

In addition, three examples of irregular stride FFTs which were misclassified as idle are shown in Figures 4.11g - 4.11i. These three figures show little activity across the frequency domain, and they do appear very similar to idle FFTs. It is likely that many of the irregular stride FFTs which are misclassified this way fall on instances where only one step is taken (perhaps even with the non-instrumented foot), so the accelerometer signal does not show any repeating patterns within a 5 second window. Because a 3 second window was determined as the boundary for an irregular stride, if one step was taken with the non-instrumented foot, it would be very easy to incorrectly classify the accelerometer signal as idle because there could actually be no activity within the accelerometer signal. A wider window size would be required to detect this type of activity.

### 4.2.3 Varying Pedometer Algorithms

If the pedometer algorithm used is selected based on the gait detected, using knowledge of sensor location, the resulting step detection accuracy can be seen in Table 4.7. This type of analysis compares the varying algorithm pedometer to the best performing individual algorithm across each metric. From this table, it can be seen that the overall step detection accuracy increases to 84%, which is 3% more accurate than the best performing single algorithm according to SDA, the peak detector. When comparing RCA, best performing individual algorithm (0.94) and the varying algorithm (1.06) provides similar accuracy measures over time. From this RCA value, it can be seen that varying the step detection algorithm would cause a small step overestimation over extended usage, but the rate of overestimation is similar to the rate of underestimation for the autocorrelation algorithm. The rate of overestimation is also significantly reduced when compared to the peak detector and threshold based algorithms.

The accuracy of varying the step detection algorithm based on gait type, with respect to sensor location, is shown in Table 4.8. This analysis is useful when comparing the varying algorithm

Algorithm	Overall RCA	Overall SDA
Peak Detector	1.64	0.81
Threshold	2.46	0.61
Autocorrelation	0.94	0.77
Varying Algorithm	1.06	0.84

Table 4.7: Step detection accuracy while varying the algorithm based upon classifier detection of gait type. When compared to the most accurate individual algorithm within each metric, RCA is the same, and SDA demonstrates a 3% improvement.

Sensor Position	Overall RCA	Overall SDA	Varying RCA	Varying SDA
Wrist	2.03	0.69	1.10	0.77
Hip	1.56	0.73	1.06	0.86
Ankle	1.46	0.76	1.02	0.82

Table 4.8: Step detection accuracy while varying algorithm based upon classifier detection of gait type, as compared to the average of three algorithms’ performances, across sensor location. All accuracies show large improvements.

pedometer against an unknown algorithm when sensor position is known, for example, comparing multiple commercial wrist worn pedometers to the varying algorithm pedometer. From this analysis, it can be seen that the step detection accuracy of the varying algorithm shows an 8% increase for a wrist worn device, 13% increase for a hip worn device, and a 6% accuracy for an ankle worn device when compared against the average of the three algorithms implemented independently. This indicates consistent improvement for the varying algorithm pedometer regardless of the position on which the sensor is worn. Since the gait detection algorithm is not created with knowledge of the sensor position, this is an impressive feat which speaks to the generalizability of the gait detection algorithm.

The accuracy of varying the step detection algorithm based on gait type, with respect to gait, is shown in Table 4.8. This analysis is useful when comparing the varying algorithm pedometer against an unknown algorithm when gait type is known (for example, the user only wears a pedometer when exercising). From this table, it can be seen that the step detection accuracy for the varying step detector is 3% greater for regular gait, 3% greater for semi-regular gait, and 19% better for unstructured gait. Based on the RCA values, the varying algorithm pedometer also generally slightly underestimates the number of steps taken for regular and semi-regular gait, but overestimates the number of steps taken during unstructured gait. The long term running count accuracy of the device in a natural environment will depend on the amount of time spent performing each activity.

Gait	Overall RCA	Overall SDA	Varying RCA	Varying SDA
Regular	0.97	0.90	0.99	0.93
Semi-regular	1.13	0.77	0.96	0.80
Unstructured	2.95	0.53	1.25	0.72

Table 4.9: Step detection accuracy while varying algorithm based upon classifier detection of gait type, as compared to the average of three algorithms’ performances, across gait type. All accuracies show large improvements.

### 4.3 Conclusions and Future Work

The goal of this chapter was to determine if gait type could be detected, and if so, to determine if pedometer algorithm performance could be improved with this knowledge. The results indicate that the correct gait type was identified with 84% accuracy. With the detected gait and knowledge of where the pedometer is worn on the body, step detection accuracy improved to 84% compared to 81% for the best performing individual algorithm. It is possible that the gait detection algorithm could be improved if the sensor location were already known.

It was also determined that in order for a gait detection algorithm to perform in real time, it should identify gait in a smaller window of time. When examining these smaller windows of time and labeling gaits, it was determined that gaits present in 30 seconds or less do not correspond directly to the names given to the activities performed in our experiment. Instead of regular gait, semi-regular gait, and unstructured gait, these smaller windows of motion were more accurately described as steady stride, irregular stride, and idle. While regular gait and steady stride are strongly related, semi-regular gait is characterized by periods of steady stride, irregular stride, and idle. Unstructured gait is also characterized by the appearance of all three stride types, although significantly more time is spent idle. A classifier was build to identify strides, and strides were correctly classified with 82% accuracy. With this window size, steady stride and idle are classified with high accuracy, but identification of irregular stride is challenging. Longer window sizes are able to more accurately identify irregular stride, but because steady stride and idle were more common throughout the activities performed in the experiment, the window able to most accurately able to identify steady stride and idle provided the highest overall accuracy.

One clear step for future work is determining pedometer algorithm accuracy at a stride level. With the knowledge of which algorithms and parameter sets lead to the best performance for specific stride types, the varying algorithm pedometer could be tested based on its ability to

provide accurate real time feedback. Another direction for future work is to develop or implement a sensor position detection algorithm with the intent of using sensor position to help improve the gait detection classifier. Because sensor position has been extensively studied in previous works, it is likely that a solution exists and could be tested on our dataset. In addition, it could be valuable to collect data from running or from other activities known to be challenging for accurate step detection like folding laundry or vacuuming.

## Chapter 5

# Conclusion

This work created a large, free, and publicly available database of accelerometer signals with the goal of standardizing the evaluation of pedometer algorithms. In particular, the work focused on gathering data from a variety of gaits seen in free living and identifying the ground truth step times through video recording. It is a goal that future publications in this area would use this data set and methods for evaluation and comparison. Prior publications use individual datasets designed to examine specific challenges in step detection. This process yields algorithms which have been tested on one dataset but rarely compared directly against other algorithms on the same dataset. This process of open, formal evaluation in the public domain could improve consumer confidence in new and prior pedometer algorithms.

There are a number of possible directions for future work. 1) It is possible that a classifier could be developed to automatically detect sensor position, using this information to select an appropriate algorithm. 2) An algorithm could be designed to detect shifts and differentiate these from steps. As a follow up to this work, the energy expenditure for a step, as compared to a shift, could be compared, potentially allowing the algorithm to more accurately estimate energy expenditure. 3) Additional published algorithms could be implemented on the dataset, providing additional opportunities for comparison. 4) As a long term goal, we seek to make this dataset publicly available, providing other researchers the opportunity to apply their algorithms and receive quantitative comparisons to other algorithms. 5) Another long term goal is to expand the dataset to consider additional gait types, including activities known to cause problems in commercial pedometers. For example, mowing the lawn, vacuuming, or riding a car along a bumpy road.

The methodology used in this work could be applied to other mHealth problems. Developing public datasets for a variety of other mHealth devices would allow algorithms for each of the devices to be directly compared. In addition, focusing on free living activities will allow the evaluations to translate well to free living environments. The critical component in evaluating free living data sets is the development of reliable ground truth information, which can be challenging. However, if participants are video recorded throughout the dataset collection, video review can allow for accurate ground truth identification. This process and methodology could lead to algorithm improvements and increases in consumer confidence across many mHealth devices.



# Appendices

## Appendix A Additional Work

The following two sections describe additional projects and publications in which this author has been involved. The first is an ongoing investigation into the source of simulator sickness produced by head mounted display (HMD) usage. The second is an extension of the work accomplished in my Master's thesis involving bite detection through use of a table embedded scale.

### A.1 HMD Latency Analysis

Head mounted displays (HMDs), virtual reality, and augmented reality are quickly becoming popular for entertainment, training, and communication. In healthcare, HMDs have been used to improve training techniques through use of virtual reality [81, 82]. One problem in the adoption of virtual reality has been the presence of simulator sickness, which can cause discomfort, apathy, drowsiness, disorientation, fatigue, and vomiting. Prior research in this area determined that latency in the tracking systems of head mounted displays (HMDs) caused this sickness [83, 84, 85]. Our group, however, examined the effects of varying latency on sickness levels and demonstrated that frequency, not amplitude correlates to increased sickness levels [86, 87]. Identifying the source of this sickness is critical to counteracting it and increasing the adoption and effectiveness of virtual reality.

This author worked with Amelia Kinsella and Eric Muth to develop software allowing for artificial latency to be added to the image displayed in a head mounted display. In particular the latency added could be varied in order to produce a sinusoidal wave at varying amplitudes and frequencies to examine the effect that this type of latency has on simulator sickness. This author also assisted in the development of software used to validate the ability of the system to accurately control the latency additions.

This author also assisted in Mike Wilson's doctoral dissertation research by developing a program in Matlab used to provide audio cues, record posture motion data, and record audio produced when buzzers were set off throughout the experiment. This program was used to run the experiment, which required participants to wear a head mounted display with live camera footage of their environment. Participants performed a task in which they activated buzzers located throughout a room through use of a laser pointer. The specific buzzer students were directed to activate was provided through audio cues, and the program identified the amount of time between the end of the

instruction and when the buzzer was activated. This time delay, in addition to the posture motion data were used to indicate performance and sickness levels of participants.

## A.2 Table Embedded Scale

My work in detecting bites using a table embedded scale was extended after the completion of my master's thesis. The work was expanded by updating the bite detection algorithm, and publishing a journal article based on my findings. In addition, statistical analyses were performed on the collected data, and another journal article was published detailing these analyses and findings. The abstracts for the two publications are given below.

The universal eating monitor (UEM) is a table-embedded scale used to measure grams consumed over time while a person eats. It has been used in laboratory settings to test the effects of anorectic drugs and behavior manipulations such as slowing eating, and to study relationships between demographics and body weight. However, its use requires restricted conditions on the foods consumed and behaviors allowed during eating in order to simplify analysis of the scale data. Individual bites can only be measured when the only interaction with the scale is to carefully remove a single bite of food, consume it fully, and wait a minimum amount of time before the next bite. Other interactions are prohibited such as stirring and manipulating foods, retrieving or placing napkins or utensils on the scale, and in general anything that would change the scale weight that was not related to the consumption of an individual bite. This paper describes a new algorithm that can detect and measure the weight of individual bites consumed during unrestricted eating. The algorithm works by identifying time periods when the scale weight is stable, and then, analyzing the surrounding weight changes. The series of preceding and succeeding weight changes is compared against patterns for single food bites, food mass bites, and drink bites to determine if a scale interaction is due to a bite or some other activity. The method was tested on 271 subjects, each eating a single meal in a cafeteria setting. A total of 24 101 bites were manually annotated in synchronized videos to establish ground truth as to the true, false, and missed detections of bites. Our algorithm correctly detected and weighed approximately 39% of bites with approximately one false positive (FP) per ten actual bites. The improvement compared to the UEM is approximately three times the number of true detections and a 90% reduction in the number of FPs. Finally, an analysis of bites that could not be weighed compared to those that could be weighed revealed no statistically significant difference in average weight. These results suggest that our algorithm could be used to conduct studies using

a table scale outside of laboratory or clinical settings and with unrestricted eating behaviors [4].

Our study investigated the relationship between BMI and bite size in a cafeteria setting. Two hundred and seventy one participants consumed one meal each. Participants were free to select any food provided by the cafeteria and could return for additional food as desired. Bite weights were measured with a table embedded scale. Data were analyzed with ANOVAs, regressions, Kolmogorov-Smirnov tests, and a repeated measures general linear model for quartile analysis. Obese participants were found to take larger bites than both normal ( $P=0.002$ ) and overweight participants ( $p=0.017$ ). Average bite size increased by 0.20 g per point increase in BMI. Food bites and drink bites were analyzed individually, showing 0.11 g/BMI and 0.23 g/BMI slopes, respectively. Quartiles of bites were also analyzed, and a significant interaction was found between normal and obese participants ( $p=0.034$ ) such that the lower two quartiles were similar, but the upper two quartiles showed an increase in bite size for obese participants. The source of these effects could be the result of a combination of several uncontrolled factors [5].

## Appendix B Publications

R. Mattfeld, E. Jesch, and A. Hoover. An evaluation of a commercial pedometer on multiple gait types using video for ground truth, in 2018 IEEE Biomedical and Health Informatics Conference, Las Vegas, NV, 2018. (in publication)

R. Mattfeld, E. R. Muth and A. Hoover, "Measuring the consumption of individual solid and liquid bites using a table embedded scale during unrestricted eating," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 6, pp. 1711-1718, 2017.

R. Mattfeld, E. Muth, and A. Hoover, A comparison of bite size and BMI in a cafeteria setting, in Physiology & Behavior, vol. 181, pp. 38-42, 2017.

R. Mattfeld, E. Jesch, and A. Hoover. A new dataset for evaluating pedometer performance, in 2017 IEEE International Conference on Bioinformatics and Biomedicine, Kansas City, MO, 2017.

A. Hoover, R. Mattfeld, and E. Muth. Bites as a unit of measurement, in Advances in the Assessment of Dietary Intake, CRC Press, pp. 149-162, 2017.

A. Kinsella, R. Mattfeld, E. Muth, and A. Hoover, Frequency, not amplitude, of latency, affects subjective sickness in a head-mounted display, in Aerospace Medicine and Human Performance, vol. 87, no. 7, pp. 604-609, 2016.

## Appendix C PAR-Q

Department of Electrical and Computer Engineering  
 College of Engineering, Computing and Applied Sciences  
 Riggs Hall  
 Clemson, SC 29634



### Physical Activity Readiness Questionnaire (PAR-Q)

Name (print): \_\_\_\_\_ Date: \_\_\_\_\_ DOB: \_\_\_/\_\_\_/\_\_\_ Age: \_\_\_\_\_

Physicians Name: \_\_\_\_\_ Physicians Phone #: (\_\_\_\_) \_\_\_\_\_

**For research use only:**

Height: \_\_\_\_\_ in.      Weight: \_\_\_\_\_ lbs.      BMI: \_\_\_\_\_ kg/m<sup>2</sup>

	Questions	Yes	No
1	Has your doctor ever said that you have a heart condition and that you should only perform physical activity recommended by a doctor?		
2	Do you feel pain in your chest when you perform physical activity?		
3	In the past month, have you had chest pain when you were not performing any physical activity?		
4	Do you lose your balance because of dizziness or do you ever lose your consciousness?		
5	Do you have a bone or joint problem that could be made worse by a change in your physical activity?		
6	Is your doctor currently prescribing any medication for your blood pressure or for a heart condition?		
7	Do you know of any other reason why you should not engage in physical activity?		

If you have answered "Yes" to one or more of the above questions, consult your physician before engaging in physical activity. Tell your physician which questions you answered "Yes" to. After a medical evaluation, seek advice from your physician on what type of activity is suitable for your current condition.

# Bibliography

- [1] Y. Dong, A. Hoover, J. Scisco, and E. Muth, “Detecting eating using a wrist mounted device during normal daily activities,” in *Proceedings of the 9th International Conference on Embedded Systems and Applications*, 2011.
- [2] Z. Huang, “An Assessment of the Accuracy of an Automated Bite Counting Method in a Cafeteria Setting,” Master Thesis, College of Engineering and Science, Clemson University, 2013.
- [3] Y. Shen, J. Salley, E. Muth, and A. Hoover, “Assessing the accuracy of a wrist motion tracking method for counting bites across demographic and food variables,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 599–606, 2017.
- [4] R. S. Mattfeld, E. R. Muth, and A. Hoover, “Measuring the consumption of individual solid and liquid bites using a table-embedded scale during unrestricted eating,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1711–1718, 2017.
- [5] R. S. Mattfeld, E. R. Muth, and A. Hoover, “A comparison of bite size and bmi in a cafeteria setting,” *Physiology and Behavior*, vol. 181, no. Supplement C, pp. 38 – 42, 2017.
- [6] Mobile medical applications. [Online]. Available: <https://www.fda.gov/MedicalDevices/DigitalHealth/MobileMedicalApplications/default.htm>
- [7] 500m people will be using healthcare mobile applications in 2015. [Online]. Available: <https://research2guidance.com/500m-people-will-be-using-healthcare-mobile-applications-in-2015-2/>
- [8] S. Crouter, P. Schneider, M. Karabulut, and D. Basset, “Validity of ten electronic pedometers for measuring steps, distance, and kcals,” *Medicine & Science in Sports & Exercise*, vol. 35, no. 5, p. S283, 2003.
- [9] E. Garcia, H. Ding, A. Sarela, and M. Karananithi, “Can a mobile phone be used as a pedometer in an outpatient cardiac rehabilitation program?” *The 2010 IEEE/ICME International Conference on Complex Medical Engineering*, pp. 250–253, 2010.
- [10] S. Nakae, Y. Oshima, and K. Ishii, “Accuracy of spring-levered and piezo-electric pedometers in primary school japanese children,” *Journal of Physiological Anthropology*, vol. 27, no. 5, pp. 233–239, 2008.
- [11] W. Park, V. Lee, B. Ku, and H. Tanaka, “Effect of walking speed and placement position interactions in determining the accuracy of various newer pedometers,” *Journal of Exercise Science & Fitness*, vol. 12, no. 1, pp. 31–37, 2014.
- [12] A. Singh, C. Farmer, M. V. D. Berg, M. Killington, and C. Barr, “Accuracy of the fitbit at walking speeds and cadences relevant to clinical rehabilitation populations,” *Disability and Health Journal*, vol. 9, pp. 320–323, 2016.

- [13] R. Saeedi, N. Amini, and H. Ghasemzadeh, "Patient-centric on-body sensor localization in smart health systems," *2014 48th Asilomar Conference on Signals, Systems and Computers*, pp. 2081–2085, 2014.
- [14] J. Sheu, G. Huang, W. Jheng, and C. Hsiao, "Design and implementation of a three-dimensional pedometer accumulating walking or jogging motions," *2014 International Symposium on Computer, Consumer and Control (IS3C)*, pp. 828–831, 2014.
- [15] N. Silcott, D. Bassett, D. Thompson, E. Fitzhugh, and J. Steeves, "Evaluation of the omron hj-720itc pedometer under free-living conditions," *Medicine & Science in Sports & Exercise*, vol. 43, no. 9, pp. 1791–1797, 2011.
- [16] Z. Tang, Y. Guo, and X. Chen, "Self-adaptive step counting on smartphones under unrestricted stepping modes," *2016 IEEE 40th Annual Computer Software and Applications Conference*, vol. 1, pp. 788–797, 2016.
- [17] A. Akahori, Y. Kishimoto, and K. Oguri, "Estimate activity for m-health using one three-axis accelerometer," *2006 3rd IEEE/EMBS International Summer School on Medical Devices and Biosensors*, pp. 122–125, 2006.
- [18] S. Jayalath and N. Abhayasinghe, "A gyroscopic data based pedometer algorithm," *2013 8th International Conference on Computer Science & Education (ICCSE)*, pp. 551–555, 2013.
- [19] S. Zhong, L. Wang, A. Bernardos, and M. Song, "An accurate and adaptive pedometer integrated in mobile health application," *IET International Conference on Wireless Sensor Network, 2010 (IET-WSN)*, pp. 78–83, 2010.
- [20] E. L. Melanson, J. R. Knoll, M. L. Bell, W. T. Donahoo, J. Hill, L. J. Nysse, L. Lanningham-Foster, J. C. Peters, and J. A. Levine, "Commercially available pedometers: considerations for accurate step counting," *Preventive Medicine*, vol. 39, no. 2, pp. 361 – 368, 2004.
- [21] G. C. Le Masurier and C. Tudor-Locke, "Comparison of pedometer and accelerometer accuracy under controlled conditions," *Medicine and science in sports and exercise*, vol. 35, no. 5, p. 867871, 2003.
- [22] (2011, December) Mems inertial sensor, lis344alh accelerometer. [Online]. Available: <http://www.st.com/internet/analog/product/207281.jsp>
- [23] Y. Lim, I. Brown, and J. Khoo, "An accurate and robust gyroscope-gased pedometer," *Conf Proc IEEE Eng Med Biol Soc*, pp. 4587–4590, 2008.
- [24] (2011, December) Mems inertial sensor, lpr410al gyroscope. [Online]. Available: <http://www.nutritionj.com/content/6/1/44>
- [25] D. Bravata, C. Smith-Spangler, V. Sundaram, A. Gienger, N. Lin, R. Lewis, C. Stave, I. Olkin, and J. Sirard, "Using pedometers to increase physical activity and improve health - a systematic review," *JAMA*, vol. 198, no. 19, pp. 2296–2304, 2007.
- [26] I. Janssen and A. Leblanc, "Review systematic review of the health benefits of physical activity and fitness in school-aged children and youth," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 7, no. 40, pp. 1–16, 2010.
- [27] M. Reiner, C. Niermann, D. Jekuac, and A. Woll, "Long-term health benefits of physical activity a systematic review of longitudinal studies," *BMC Public Health*, vol. 13, no. 813, pp. 1–9, 2013.
- [28] D. E. Warburton, C. W. Nicol, and S. S. Bredin, "Health benefits of physical activity: the evidence," *CMAJ*, vol. 174, no. 6, pp. 801–809, 2006.



- [29] T. Choudhury, S. Consolvo, B. Harrison, J. Hightower, A. LaMarca, L. LeGrand, A. Rahimi, A. Rea, G. Borriello, B. Hemingway, P. Klasnja, K. Koscher, J. Landay, J. Lester, D. Wyatt, and D. Haehnel, “The mobile sensing platform: An embedded activity recognition system,” *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 32–41, 2008.
- [30] E. Kim, S. Helal, and D. Cook, “Human activity recognition and pattern discovery,” *Pervasive Computing, IEEE*, vol. 9, no. 1, pp. 48–53, January–March 2010.
- [31] J. Kwapisz, G. Weiss, and S. Moore, “Activity recognition using cell phone accelerometers,” *SIGKDD Explor. Newsl.*, vol. 12, no. 2, pp. 74–82, 2011.
- [32] U. Maurer, A. Smailagic, D. Siewiorek, and M. Deisher, “Activity recognition and monitoring using multiple sensors on different body positions,” *International Workshop on Wearable and Implantable Body Sensor Networks (BSN’06)*, pp. 4–8, 2006.
- [33] T. van Kasteren, A. Noulas, G. Englebienne, and B. Kröse, “Accurate activity recognition in a home setting,” *Proceedings of the 10th International Conference on Ubiquitous Computing*, pp. 1–9, 2008.
- [34] G. Welk, S. Blair, K. Wood, S. Jones, and R. Thompson, “A comparative evaluation of three accelerometry-based physical activity monitors,” *Medicine and Science in Sports and Exercise*, vol. 32, pp. S489–497, 2000.
- [35] Y. Dong, “Tracking Wrist motion to Detect and Measure the Eating Intake of Free-Living Humans,” PhD dissertation, Electrical and Computer Engineering Department, Clemson University, SC., 2012.
- [36] A. Smeets and M. Westerterp-Plantenga, “Acute effects on metabolism and appetite profile of one meal difference in the lower range of meal frequency,” *British Journal of Nutrition*, vol. 99, pp. 1316–1321, 5 2008.
- [37] J. Speakman, *Doubly Labelled Water: Theory and Practice*. Springer, 1997.
- [38] G. Plasqui, A. Joosen, A. Kester, A. Goris, and K. Westerterp, “Measuring free-living energy expenditure and physical activity with triaxial accelerometry,” *Obesity Research*, vol. 13, no. 8, pp. 1363–1369, 2005.
- [39] D. Schoeller *et al.*, “Measurement of energy expenditure in free-living humans by using doubly labeled water,” *The Journal of Nutrition*, vol. 118, no. 11, pp. 1278–1289, 1988.
- [40] J. Noah, D. Spierer, J. Gu, and S. Bronner, “Comparison of steps and energy expenditure assessment in adults of fitbit tracker and ultra to the actual and indirect calorimetry,” *Journal of Medical Engineering & Technology*, vol. 37, no. 7, pp. 456–462, 2013.
- [41] L. Lanningham-Foster, R. Foster, S. McCrady, T. Jensen, N. Mitre, and J. Levine, “Activity-promoting video games and increased energy expenditure,” *The Journal of Pediatrics*, vol. 154, no. 6, pp. 819–823, 2009.
- [42] W. Leonard, “Laboratory and field methods for measuring human energy expenditure,” *American journal of human biology : the official journal of the Human Biology Council*, vol. 24, pp. 372–384, 2012.
- [43] K. A. Croteau, “A preliminary study on the impact of a pedometer-based intervention on daily steps,” *American Journal of Health Promotion*, vol. 18, no. 3, pp. 217–220, 2004.

- [44] S. Mansi, S. Milosavljevic, G. Baxter, S. Tumilty, and P. Hendrick, "A systematic review of studies using pedometers as an intervention for musculoskeletal diseases," *BMC Musculoskeletal Disorders*, vol. 15, no. 231, 2014.
- [45] C. Tudor-Locke, J. Williams, J. Reis, and D. Pluto, "Utility of pedometers for assessing physical activity," *Sports Medicine*, vol. 32, no. 12, pp. 795–808, 2002.
- [46] H. Montoye, H. Kemper, W. Saris, and R. Washburn, "Measuring physical activity and energy expenditure," *Human Kinetics*, vol. 68, no. 6, pp. 72–79, 1996.
- [47] J. Levine, P. Baukol, and K. Westerterp, "Validation of the tracmor triaxial accelerometer system for walking," *Med Sci Sports Exerc*, vol. 33, pp. 1593–1597, 2001.
- [48] F. Gu, K. Khoshelham, J. Shang, F. Yu, and Z. Wei, "Robust and accurate smartphone-based step counting for indoor localization," *IEEE Sensors Journal*, vol. 17, no. 11, pp. 3453–3460, 2017.
- [49] Y. Cho, H. Cho, and C. Kyung, "Design and implementation of practical step detection algorithm for wrist-worn devices," *IEEE Sensors Journal*, vol. PP, no. 99, pp. 1–1, 2016.
- [50] B. Liu, D. Wang, S. Li, X. Nie, S. Xu, B. Jiao, X. Duan, and A. Huang, "Design and implementation of an intelligent belt system using accelerometer," *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2043–2046, 2015.
- [51] J. Chien, K. Hirakawa, J. Shieh, H. Guo, and Y. Hsieh, "An effective algorithm for dynamic pedometer calculation," *2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, pp. 366–368, 2015.
- [52] K. Ozcan and S. Velipasalar, "Robust and reliable step counting by mobile phone cameras," *Proceedings of the 9th International Conference on Distributed Smart Cameras*, pp. 164–169, 2015.
- [53] H. Lee, S. Choi, and M. Lee, "Step detection robust against the dynamics of smartphones," *Sensors*, vol. 15, no. 10, pp. 27 230–27 250, 2015.
- [54] H. Montoye, H. Kemper, W. Saris, and R. Washburn, "Apfiloc: An infrastructure-free indoor localization method fusing smartphone inertial sensors, landmarks and map information," *Sensors*, vol. 15, no. 10, pp. 27 251–27 272, 2015.
- [55] M. Pan and H. Lin, "A step counting algorithm for smartphone users: Design and implementation," *IEEE Sensors Journal*, vol. 15, no. 4, pp. 2296–2305, 2015.
- [56] A. Lin, J. Zhang, K. Lu, and W. Zhang, "An efficient outdoor localization method for smartphones," *2014 23rd International Conference on Computer Communication and Networks (ICCCN)*, pp. 1–8, 2014.
- [57] M. Susi, V. Renaudin, and G. Lachapelle, "Motion mode recognition and step detection algorithms for mobile phone users," *Sensors*, vol. 13, no. 2, pp. 1539–1562, 2013.
- [58] A. Brajdic and R. Harle, "Walk detection and step counting on unconstrained smartphones," *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 225–234, 2013.
- [59] M. Oner, J. Pulcifer-Stump, P. Seeling, and T. Kaya, "Towards the run and walk activity classification through step detection - an android application," *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1980–1983, 2012.

- [60] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. Choudhury, "No need to war-drive: Unsupervised indoor localization," *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, pp. 197–210, 2012.
- [61] A. Rai, K. Chintalapudi, V. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," *MobiCom12*, pp. 293–304, 2012.
- [62] S. Li, Z. Ling, J. Cao, K. Li, and G. Liu, "A step detection algorithm based-on chain code," *2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN)*, pp. 164–167, 2011.
- [63] C. Lo, C. Chiu, Y. Tseng, S. Chang, and L. Kuo, "A walking velocity update technique for pedestrian dead-reckoning applications," *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1249–1253, 2011.
- [64] J. Chon and H. CHa, "Lifemap: A smartphone-based context provider for location-based services," *IEEE Pervas. Comput.*, vol. 10, no. 2, pp. 58–67, 2011.
- [65] P. Goyal, V. J. Ribeiro, H. Saran, and A. Kumar, "Strap-down pedestrian dead-reckoning system," *2011 International Conference on Indoor Positioning and Indoor Navigation*, pp. 1–7, 2011.
- [66] O. Bebek, M. Suster, S. Rajgopal, M. Fu, X. Huang, M. Cavusoglu, D. Young, M. Mehregany, A. van den Bogert, and C. Mastrangelo, "Personal navigation via shoe mounted inertial measurement units," *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1052–1058, 2010.
- [67] N. Zhao, "Full-featured pedometer design realized with 3-axis digital accelerometer," *Analog Dialogue*, vol. 44, no. 6, pp. 1–5, 2010.
- [68] M. Mladenov and M. Mock, "A step counter service for java-enabled devices using a built-in accelerometer," *Proceedings of the 1st International Workshop on Context-Aware Middleware and Services: Affiliated with the 4th International Conference on Communication System Software and Middleware (COMSWARE 2009)*, pp. 1–5, 2009.
- [69] M. Marschollek, M. Goevercin, K. H. Wolf, B. Song, M. Gietzelt, R. Haux, and E. Steinhagen-Thiessen, "A performance comparison of accelerometry-based step detection algorithms on a large, non-laboratory sample of healthy and mobility-impaired persons," *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1319–1322, 2008.
- [70] J. Ojeda and J. Borenstein, "Non-gps navigation with the personal dead-reckoning system," *SPIE Defense and Security Conference*, pp. 1–11, 2007.
- [71] H. Jang, J. Kim, and D. Hwang, "Robust step detection method for pedestrian navigation systems," *Electronics Letters*, vol. 43, no. 14, 2007.
- [72] S. Beauregard, "A helmet-mounted pedestrian dead reckoning system," *IFAWC 06*, pp. 1–11, 2006.
- [73] N. Ichinoseki-Sekine, Y. Kuwae, Y. Higashi, T. Fujimoto, M. Sekine, and T. Tamura, "Improving the accuracy of pedometer used by the elderly with the fft algorithm," *Medicine and Science in Sports and Exercise*, vol. 38, no. 9, pp. 1674–1681, 2006.
- [74] J. Kim, H. Jang, D. Hwang, and C. Park, "A step, stride and heading determination for the pedestrian navigation system," *Journal of Global Positioning Systems*, vol. 3, no. 1-2, pp. 273–279, 2004.

- [75] J. Lu, T. Zhang, F. Hu, Y. Wu, and K. Bao, “Measuring activities and counting steps with the smartsocks - an unobtrusive and accurate method,” *2014 IEEE Global Humanitarian Technology Conference (GHTC)*, pp. 694–698, 2014.
- [76] R. Mattfeld, E. Jesch, and A. Hoover, “A new dataset for evaluating pedometer performance,” *2017 IEEE International Conference on Bioinformatics and Biomedicine*.
- [77] D. Chisholm, M. Collis, L. Kulak, W. Davenport, and N. Gruber, “Physical activity readiness,” *British Columbia Medical Journal*, vol. 17, pp. 375–378, 1975.
- [78] How accurate are fitbit devices? [Online]. Available: [https://help.fitbit.com/articles/en\\_US/Help\\_article/1136](https://help.fitbit.com/articles/en_US/Help_article/1136)
- [79] How does my fitbit device count steps? [Online]. Available: [https://help.fitbit.com/articles/en\\_US/Help\\_article/1143](https://help.fitbit.com/articles/en_US/Help_article/1143)
- [80] J.-S. Wang and F.-C. Chuang, “An accelerometer-based digital pen with a trajectory recognition algorithm for handwritten digit and gesture recognition,” *IEEE Transactions on Industrial Electronics*, vol. 59, no. 7, pp. 2998–3007, July 2012.
- [81] S. Adamovich, G. Fluet, E. Tunik, and A. Merians, “Sensorimotor training in virtual reality: A review,” *Neurorehabilitation*, vol. 25, no. 1, pp. 29–44, 2009.
- [82] M. Graafland, J. Schraagen, and M. Schijven, “Systematic review of validity of serious games for medical education and surgical skills training,” *British Journal of Surgery*, vol. 99, no. 10, pp. 1322–1330, 2012.
- [83] P. DiZio and J. Lackner, “Circumventing side effects of immersive virtual environments,” *HCI*, pp. 893–896, 1997.
- [84] S. Jennings, G. Craig, L. Reid, and R. Kruk, “The effect of visual system time delay on helicopter control,” *Human Factors*, vol. 44, no. 13, pp. 69–72, 2000.
- [85] S. Jennings, L. Reid, G. Craig, and R. Kruk, “Time delays in visually coupled systems during flight test and simulation,” *Journal of Aircraft*, vol. 41, no. 6, pp. 1327–1335, 2004.
- [86] M. St.Pierre, S. Banerjee, A. Hoover, and E. Muth, “The effects of 0.2 hz varying latency with 20100ms varying amplitude on simulator sickness in a helmet mounted display,” *Displays*, vol. 36, pp. 1–8, 2015.
- [87] A. Kinsella, R. Mattfeld, E. Muth, and A. Hoover, “Frequency, not amplitude, of latency affects subjective sickness in a head-mounted display,” *Aerospace Medicine and Human Performance*, vol. 87, no. 7, pp. 604–609, 2016.