UCLA
UCLA Electronic Theses and Dissertations

Title
Design and Development of Performance Metrics for Elite Runners

Permalink
https://escholarship.org/uc/item/9pc8798h

Author
Mittal, Nikhil R.

Publication Date
2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Design and Development of Performance Metrics for Elite Runners

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Electrical Engineering

By

Nikhil R Mittal

2012

ABSTRACT OF THE THESIS

Design and Development of Performance Metrics for Elite Runners

By

Nikhil R Mittal

Master of Science in Electrical Engineering

University of California, Los Angeles, 2012

Professor William J. Kaiser, Chair

Recent advancements in mobile health applications of human motion sensing systems have enabled the proliferation of low-cost solutions in health and athletics assessment. As sensor data analysis and systems become more sophisticated the next innovation in this field is to provide user guidance towards improved outcomes. To this end, physiology research in distance running provides the framework to compute physics-based metrics to characterize running performance. This thesis presents a novel running gait cycle characterization system to monitor the race performance of elite runners based on several monotonic, computationally light metrics and an overall efficiency metric. The presented real time race results demonstrate the ability of motion sensing algorithms and systems to monitor, guide and enhance runner performance.

The thesis of Nikhil R Mittal is approved.

Robert N. Candler

Gregory J. Pottie

William J. Kaiser, Committee Chair

University of California, Los Angeles

2012

TABLE OF CONTENTS

1 INTRODUCTION 1

1.1 Overview 1

1.2 Motivation and Background 2

1.3 Thesis Contribution 4

2 SENSOR SYSTEMS 7

2.1 Basic Sensor Principles 7

2.2 Accelerometers 8

2.3 Data Aggregation & Analysis 10

2.3.1 Sensor Node 12

2.3.2 Online Data Analysis Node 13

2.3.3 Offline Data Analysis Network 15

2.4 Prototype Running Performance System: MDAWN Datalogger 16

2.4.1 Specification 16

2.4.2 Data Management 19

2.4.3 Sensor Placement 20

2.4.4 Sensing System Orientations 21

3 PERFORMANCE METRICS 22

3.1 Metric characteristics 22

3.2 Metric description 24
3.2.1 Cadence 25
3.2.2 Ground Contact Time (GCT) 25
3.2.3 Distal Leg Lift (DLL) 31
3.2.4 Over-stride Metric 34
3.2.5 Contact De-acceleration Energy Loss (CDEL) Metric 35
3.2.6 Pronation 37
4 ALGORITHM ARCHITECTURE AND STRIDE SEGMENTATION 41
4.1 Pre-processing of raw data 42
4.1.1 Stride segmentation/recognition 42
4.1.2 Activity Detection 45
4.2 Real time metric computation 46
4.3 Post processing of data 46
5 REAL TIME DEPLOYMENT AND RACE RESULTS 48
5.1 Real Time Deployment 49
5.2 Race Results 50
5.3 Los Angeles Rock & Roll Half Marathon 51
5.3.1 Final Race Results 52
5.3.2 Cadence 52

5.3.3 Ground Contact Time (GCT) 54
5.3.4 Distal Leg Lift (DLL) 57
5.3.5 Over-stride Metric 59
5.3.6 Contact De-acceleration Energy Loss (CDEL) 62
5.3.7 Overall Efficiency Metric 64
5.4 Honolulu Marathon 65
5.4.1 Final Race Results 66
5.4.2 Cadence 68
5.4.3 Ground Contact Time (GCT) 69
5.4.4 Distal Leg Lift (DLL) 72
5.4.5 Over-stride Metric 74
5.4.6 Contact De-acceleration Energy Loss (CDEL) 76
5.4.7 Overall Efficiency Metric 78
5.5 Other Important Results 79
5.5.1 Angela Bizzari 80
5.5.2 Morgan Uceny 81
5.5.3 Anna Pierce 83
5.5.4 Key Observations 84
6 CONLUSIONS AND FUTURE WORK 86

Bibliography..... 89

LIST OF FIGURES

Figure 2.1: High-level sensor diagram 7

Figure 2.2: Example Distributed wireless sensing system architecture 11

Figure 2.3: High-level hardware architecture of a sensor node 12

Figure 2.4: MDAWN Data-logger prototype 17

Figure 2.5: Sparkfun Electronics Logomatic v2 17

Figure 2.6: MDAWN Data uploader software interface 19

Figure 2.7: MDAWN sensor placement 21

Figure 2.8: Sensor system orientation 21

Figure 3.1: Typical human running gate cycle 24

Figure 3.2: Ground Contact phase of a rear foot striker (RFS) 27

Figure 3.3: Typical z-axis data 28

Figure 3.4: Graph highlighting the monotonic nature of average GCT_F values for the five cases mentioned in Table 3.1 30

Figure 3.5: DLL metric vs. the increasing degree of leg lift 33

Figure 3.6: Over-stride metric vs. increasing degree of stride length 35

Figure 3.7: System Acceleration in Abduction Direction for Left Shoe in Tempo Run 39

Figure 3.8: Combined Left and Right Foot Pronation Rotation at different stages of the run 40

Figure 4.1: Work flow for algorithm design 41

Figure 4.2: Raw acceleration data from the tri-axial accelerometer 42

Figure 4.3: Data vector generated for stride segmentation purposes highlighting the margin of error 43

Figure 5.1: A screenshot of the real time display 50

Figure 5.2: Course Elevation chart for the LA R&R Half Marathon 51

Figure 5.3: Final race timings and comparison of the three runners 52

Figure 5.4: Distance vs. Cadence time series for the three runners 53

Figure 5.5: GCT_S of right feet vs. Distance for the three runners 54

Figure 5.6: GCT_F of right feet vs. Distance for the three runners 55

Figure 5.7: GCT_S of left feet vs. Distance for the three runners 55

Figure 5.8: GCT_F of left feet vs. Distance for the three runners 56

Figure 1.9: Distal Leg Lift metric with distance for Josh for both feet 57

Figure 5.10: Distal Leg Lift Metric with distance for John for both feet 58

Figure 5.11: Distal Leg Lift metric with distance for Brian for both feet 58

Figure 5.12: Over-stride metric vs. Distance for Josh 60

Figure 5.13: Over-stride metric vs. Distance for John 60

Figure 5.14: Over-stride metric vs. Distance for Brian 61

Figure 5.15: CDEL vs. Distance for Josh 62

Figure 5.16: CDEL vs. Distance for John 63

Figure 5.17: CDEL vs. Distance for Brian 63

Figure 5.18: Overall Efficiency metric for the three runners 65

Figure 5.19: Course Elevation chart for the Honolulu Marathon 66

Figure 5.20: Final race timings and comparison of the three runners 67

Figure 5.21: Split timing comparison for the three runners 68

Figure 5.22: Distance vs. Cadence time series for the three runners 68

Figure 5.23: GCT_S of right feet vs. Distance for the three runners 70

Figure 5.24: GCT_F of right feet vs. Distance for the three runners 70

Figure 5.25: GCT_S of left feet vs. Distance for the three runners 71

Figure 5.26: GCT_F of left feet vs. Distance for the three runners 71

Figure 5.27: Distal Leg Lift metric with distance for Josphat for both feet 73

Figure 5.28: Distal Leg Lift metric with distance for Jimmy for both feet 73

Figure 5.29: Distal Leg Lift metric with distance for Patrick for both feet 74

Figure 5.30: Over-stride metric vs. Distance for Josphat 75

Figure 5.31: Over-stride metric vs. Distance for Jimmy 75

Figure 5.32: Over-stride metric vs. Distance for Patrick 76

Figure 5.33: CDEL metric vs. Distance for Josphat 77

Figure 5.34: CDEL metric vs. Distance for Jimmy 77

Figure 5.35: CDEL metric vs. Distance for Patrick 78

Figure 5.36: Overall Efficiency metric with Distance for the three runners 79

Figure 5.37: Progression of GCT_S for the six 1K runs by Angela 80

Figure 5.38: Progression of DLL metric for the six 1K runs by Angela 81

Figure 5.39: Progression of GCT_S for the different runs by Morgan 82

Figure 5.40: Progression of DLL metric for the different runs by Morgan 82

Figure 5.41: Progression of GCT_S for the different runs by Anna 83

Figure 5.42: Progression of DLL metric for the different runs by Anna 84

LIST OF TABLES

Table 2.1: Additional sensor figures of merit 8

Table 2.2: Figures of merit for Analog Devices ADXL326 accelerometer [17] 10

Table 2.3: Typical sensor node system components 13

Table 2.4: Motorola droid bionic specifications [19] 15

Table 2.5: lPC2148 main specifications [20] 17

Table 3.1: GCT_F values for runs with different technical flaws 29

Table 3.2: DLL metric values for different types of runs for the 3 runners 33

Table 3.3: Average CDEL metric values for 5 normal and 5 heavy footed runs 37

ACKNOWLEDGEMENTS

This thesis is the result of my work here at UCLA in my capacity as a graduate student. I feel deeply indebted to the different sources of support I've received in my time here at UCLA. This work would not have been possible without my faculty adviser Professor William J. Kaiser. He has been a source of inspiration and guidance throughout. Every meeting with him is an important lesson in critical thinking and result oriented analysis. I would also like to thank Professor Greg Pottie and Professor Robert Candler for agreeing to serve on my thesis committee.

My peers on the Pegasus team deserve a special mention for their support. Specifically, Yeung Lam, Mahdi Ashktorab and John Larson whose own efforts on the project have made this work possible. I would also like to acknowledge other colleagues in the ASCENT lab including Yan Wang, Celia Xu and Victor Chen who have taken time out of their busy schedules to impart their knowledge and experience.

Finally, I would like to extend my deepest gratitude to my family and friends (especially Gaelen Pereira, Saad Ahmad and Ishita Rajani) for their support and understanding.

1 INTRODUCTION

1.1 Overview

Recent developments in the field of embedded systems, due to the continuous decrease in the size of electronic devices, have enabled increased device mobility. With increased mobility, these devices are deployed in innovative applications which were unheard of before. A burgeoning field of mobile electronic devices is the wireless sensing system industry. New capabilities and ease of availability of these sensor devices enable, through distributed communications networks, collection of large amounts of data to be processed in order to address application specific requirements. Applications may range from wireless health care to energy monitoring and performance characterization, each presenting its own unique set of system design challenges.

In the world of competitive running, professional runners strive to find extra pace every day. Running pace, especially over long distances, is not only about endurance and effort but also relies heavily on the running technique. In the past runners have solely relied on coaches to give them feedback on their technique and their faults. While being effective, it is a matter of perception and more of subjective art rather than an objective analysis.

Injury prevention and staying fit is a major area of concern for every professional athlete. The running technique of a runner is a major determining factor in how much prone to injury a particular runner is. Most elite distance runners take, on average, between 180 and 200 foot strikes per minute of racing. Over the course of a 10,000m race, elite males will take approximately 5100 steps and elite females approximately 5800 steps.

The mere repetitive nature of this activity highlights the importance of analyzing and understanding each runner's technique and modifying it if necessary to reduce the injury risks.

In the last decade, there has been revolutionary development in the field of sensor technology. There are now commercial products available in the markets which give real time pace and the distance covered. Most products, although immensely helpful, do not provide any insight into the running technique or gait analysis. Using sensors it is now possible to analyze the gait cycle of individual runners.

Many races competed at the elite level are decided by a margin of seconds. Hence, every runner wants to optimize his/her performance. Analysis of their runner technique provides them with a unique insight which can prove to be highly useful in performance improvement. It is our endeavour to analyze the biomechanical features of a runner's stride in real time and help them improve their running biomechanical efficiency.

1.2 Motivation and Background

Running is a sport that appeals to everyone and yet little information is available as to what constitutes a good way to run. While being a leisure and fitness activity for most, it is also a fiercely contested sport. Despite its popularity, running is considered to be a high injury risk sport with 65 % of runners being injured in an average year [1]. There is much debate going on as to what should be the ideal physical attributes that contribute to a perfect running stride. The answer to that question might be subjective and vary from individual to individual; running performance has been modeled as a function of a lot of parameters highly dependent on the individual runner. While it's agreed upon that technique should

rather be analyzed and optimized for the individual runner and her individual condition [2], we attempt to model some characteristic features of a running stride that contribute significantly to the running performance and help in injury prevention.

Existing analyses mainly base on motion data collected with optical motion capture systems (e.g. Vicon) and ground reaction force measurements [3]. These studies are conducted in controlled conditions using means like force platforms, measurement of oxygen uptake and video analysis techniques [4]. While these systems provide high measurement accuracy, they are restricted to laboratory environments and do not allow for continuous monitoring in the field. This has been a major source of hindrance in real time race monitoring. With the technology now at our disposal, we attempt to make use of data analysis techniques on data from tri-axial accelerometers worn at the foot of the runners to characterize the gait cycle of a runner.

The systems used for gait analysis until now, mentioned above, have limited use also because of the cost and computational complexity for data analysis of such systems. This restricts their use to research and to elite athletes [5]. There is thus a need for a cheap, affordable and computationally efficient product that performs gait analysis. We will present such a system here which will not only have performance measurement capabilities but will also have the potential to act as a guidance tool for better and focused training.

Currently some commercial products are available catering to the needs of the running community. None of them, however, have the sophistication required to carry out a comprehensive gait analysis. Analysis capabilities of current system such as the Nike+ foot sensor [6] and Polar RS-100 training computer [7] are limited to measurement of pace,

distance and calories burned. More specifically, running performance characterization systems do not yet provide guidance to the athlete to enhance performance and avoid injuries that could hinder training.

Running technique of a runner evolves as the athlete reaches his/her peak. In fact, owing to exhaustion and possibly other factors, the gait cycle characteristics change over the course of a race. It has been found that fatigue alters the running kinematics, which increases the risk of injury [8], [9], [10] and affects the running economy [11], [12]. However, the complex relationship between running kinematics, injury, and running economy is still not very well understood [11]. As we will observe later, the effects of fatigue are pronounced and can be easily observed using the algorithms developed in this thesis.

1.3 Thesis Contribution

Having established the pressing need for such real time analytical capabilities, it is important here to mention that this work was done with a very precise mission in mind.

Mission Statement: Deliver the first direct, in-field monitoring and guidance system advancing running efficiency and injury risk reduction

Keeping in mind the above aim, the following approach was adopted in order to arrive at a solution to this challenging problem:

Fundamental Approach:

1. Identify the gait motion behaviors that clinical trials have shown to be direct determinants of running efficiency

2. Use sensor data to monotonically represent the gait motion behavior in question i.e. development of metrics representative of running efficiency
3. Enable real time deployment of these metrics using computationally inexpensive algorithms

This work is aimed at the development of human motion sensing systems in running performance applications. Following are the main contributions of this thesis:

1. Physics-based metrics that are shown to provide a monotonically responsive measure of each of the critical gait motion behaviors
2. Computationally efficient algorithm implementations designed for operation on Android platforms
3. A running performance characterization system to monitor the race performance of runners i.e. running efficiency with potentially real time predictive capabilities

It is worth mentioning here that to enable and implement the above, parallel work has been done by other members of the team to develop:

- Low cost, integrated, wireless tri-axial accelerometer sensors compatible with training and competition shoes
- An Android smart phone real time display and data transport gateway
- A video analysis tool for gait analysis [13]

It is also important to mention that this thesis focuses on the lower body movements, specifically the foot. It is well established that upper body motions also have a significant impact on the running economy. This work, however, restricts its scope to analysis of lower body movements.

The thesis is organized as follows. Section 2 provides an overview of wireless sensing system design principles and describes the new system prototype. Section 3 describes key performance metrics, their relevance and the development of motion sensing algorithms to characterize the running gait cycle. Section 4 describes the resulting algorithm architecture for the real time system and discusses the stride recognition algorithm in detail. Section 5 provides key results and observations drawn from the data analysis of six elite runners in competitive race conditions and three elite runners in training sessions. Section 6 summarizes key conclusions of this thesis and details options for future work.

2 SENSOR SYSTEMS

This section will focus on low-level and high-level technical specifications of systems that are designed to solve the problems faced designing a sensor node to meet functional requirements. The focus will be on wireless sensing systems for human motion monitoring, especially as this applies to running performance analysis.

2.1 Basic Sensor Principles

Abstractly, a sensor can be considered a black box with a physical signal as an input and an electronic signal, analog or digital, as an output. Many modern sensing systems will contain a microprocessor to use programs to perform data analysis on digital signals. The microprocessor itself may perform the analog to digital conversion.

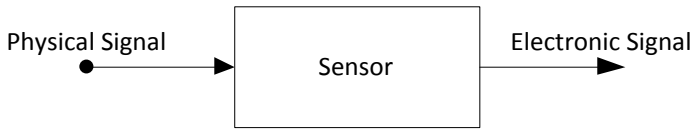


Figure 2.1: High-level sensor diagram

Ideally, the response of the sensor to the physical signal is linear. However there are numerous sources of noise and error that occur within the system that can cause the response to be non-ideal. For this reason it is necessary to define figures of merit that can characterize a sensor and help guide the design process. The first of these figures of merit is the responsivity, R, the ratio of the electronic output signal amplitude, A_{out}, to the physical signal input amplitude, A_{in} [14].

Responsivity defines the transfer function of the sensor for circuit design; it does not measure performance. The second figure of merit is sensitivity, S, the ratio of the

electronic output signal amplitude when there is no input signal i.e. A_{in} = 0 to the responsivity [14].

Sensitivity can be interpreted as a measure of the noise of the sensor due to the zero input quality. Noise is not constant in time and therefore has an associated frequency spectrum. Therefore sensitivity is measured for a wide spectrum of frequencies. Other figures of merit worth noting are presented in Table 2.1 though these are not always defined for every type of sensor.

Figure of Merit	**Definition**
Repeatability	The variation of the sensor output when it is tested under the same input conditions multiple times.
Linearity (Error)	The sensor's responsivity over the range of rated inputs of the sensor.
Drift	The change in sensor output for a constant input.
Hysteresis	The dependence of the sensor output on the dynamic conditions of the input. It is often thought of as sensor "memory" in that it describes the dependence of the output not only on the current state but also past states.
Response Time	The delay in the sensor output to respond to an input.

Table 2.1: Additional sensor figures of merit

2.2 Accelerometers

Based on the type of the physical input, a sensor can be classified into various categories. For the application of measuring running performance, we need sensors to transform motion into an electrical signal i.e. motion sensors. This class of sensors includes accelerometers for measuring acceleration, gyroscopes for measuring angular velocity and

magnetometers which measure the orientation of the sensor with regard to the Earth's magnetic field. The focus here will be on accelerometers.

Accelerometers are implemented as a micro electro-mechanical system (MEMS) that may be thought of conceptually as a proof mass on a damped spring. When the proof mass is accelerated the spring is deflected and the resulting displacement is measured. A common method to measure this displacement is capacitive sensing due to its ease of manufacturing in MEMS devices since it requires little additional processing [15]. In this implementation the proof mass is suspended with perpendicular fingers extending from the mass parallel to fixed probes which measure a differential capacitance resulting from the proof mass deflection. These sensing probes may be laid out in what is commonly referred to as a "comb" where multiple sensing capacitors are arranged parallel to each other. This MEMS structure has fabrication methods which offer tri-axial capacitive accelerometers [16].

There are several important design considerations in specific regard to use of an accelerometer in a sensing system. First, the existence of an inertial reference frame is always present and so acceleration is measured relative to this frame. In example, an accelerometer at rest relative to the Earth's surface would measure 1g upwards since any object at rest is counteracting the 1g downward pull of the Earth's acceleration. Second, being that the system is composed of a proof mass on a spring with damping there is a resonant frequency associated with the system. For acceleration frequencies much less than the resonant frequency this is not a concern however at high frequencies the dynamic-response of the sensor decays [14] which results in an operational bandwidth of the

sensor. Lastly, since accelerometers measure motion in multiple axes there are multiple sensor structures that may interfere with each other, also known as cross-talk.

Figure of Merit	Value
Range	±16 g
Linearity	±0.3%
Cross Axis Sensitivity	±1%
Sensitivity at Outputs	57 mV/g
Zero-g Bias (X_{OUT}, Y_{OUT} & Z_{OUT})	1.5 V (V_S = 3 V)
Bandwidth (X_{OUT} & Y_{OUT})	1600 Hz
Bandwidth (Z_{OUT})	550 Hz
Sensor Resonant Frequency	5.5 kHz
Supply Voltage (V_S)	1.8 – 3.6 V

Table 2.2: Figures of merit for Analog Devices ADXL326 accelerometer [17]

To illustrate these principles, Table 2.2 provides a summary of the figures of merit for a tri-axial capacitive accelerometer sensor with a measurement range well suited to the running performance characterization, the Analog Devices ADXL326. There are a few important points to note that affect the design of the wireless sensing system that would incorporate this sensor. The zero-g bias implies that the output is an analog output since the value is in volts. Other accelerometers output digital measurements using a protocol such as I^2C however this sensor outputs three analog voltages, one for each axis, so an A/D conversion stage is necessary.

2.3 Data Aggregation & Analysis

Sensors provide measurement data which must be processed to accomplish the system's stated functionality. This processing can often be computationally intensive analysis on a large set of data. In order to conserve system resources, such as battery power and memory, computation may be off-loaded to remote data-analysis nodes in a practice known as system partitioning [18]. In the case of a partitioned, or distributed, sensor system the

design of data transport from the sensor node to a data-analysis node presents many important design considerations.

Data analysis may be required in real-time, known as online analysis, to provide a user with feedback so the user modifies behavior accordingly. In other cases data analysis may be performed at a later time, known as offline analysis. If the system performs offline data-analysis, the sensor node can be out of range of the data-analysis node, store data itself and upload it at a later time when it is able to communicate with the data-analysis node. For an online data analysis system this is simply not possible. Figure 2.2 depicts the system with connectors between components indicating the transfer of data.

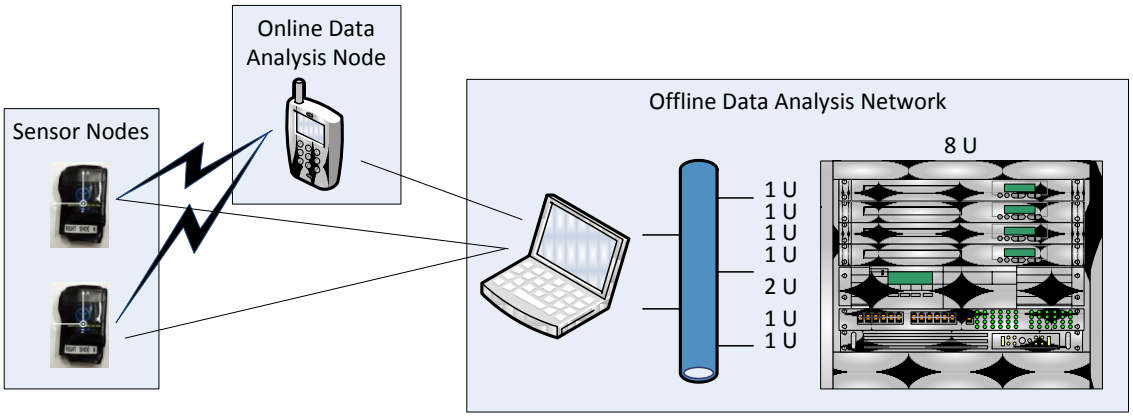


Figure 2.2: Example Distributed wireless sensing system architecture

The sensor nodes are responsible for collecting sensor data. The data may be stored or transmitted wirelessly for real time analysis to the online data analysis node. Some preprocessing of data may occur prior to storage or transmission to reduce the size of the stored data. The online data analysis node is demonstrated by a Smartphone in this example. It has the advantage of greater computational power than the sensor nodes and a screen for displaying real-time data while still being mobile. The real-time analysis node

may also be implemented to store the raw sensor data it receives for transfer to the offline data analysis network. Lastly, data can be uploaded to the offline data analysis network for more complex processing. The transfer of data and its analysis defines how each individual system component achieves its functionality.

It is important to study each of the three main components of the system described above individually in detail.

2.3.1 Sensor Node

Within wireless sensor systems the sensor node is the subject of much of the design process. In order for the system design to meet the measurement requirements the sensor node needs to be designed in both hardware and software. In comparison, the data-analysis nodes are often pre-built hardware packages such as a Smartphone or server requiring only application development on top of the preexisting systems.

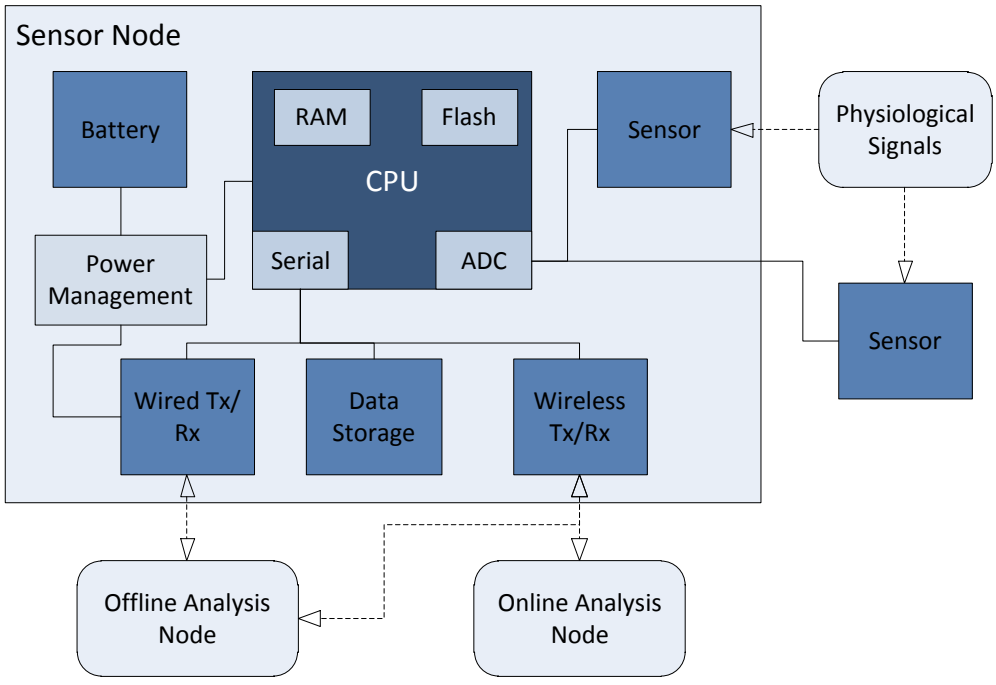


Figure 2.3: High-level hardware architecture of a sensor node

Figure 2.3 provides example high-level hardware architecture for a sensor node and Table 2.3 provides brief descriptions of the main hardware components. The CPU is typically either a microprocessor or microcontroller and performs the core functionality of the sensor node through programming. Flash memory can be programmed with boot instructions as well as program code and the RAM can have code dynamically allocated to it as well though this will not persist across reboots. The ADC is included on-board the CPU and is responsible for taking analog sensor signals and converting it to a digital value that the CPU can interpret. As discussed earlier sensors may have a digital output as well in which case they might be connected to a serial interface such as I^2C. Once sensor data is received the CPU may process it such as performing compression or feature extraction. This data can then in turn be sent over the serial interface to one of the data management devices. Here there is only one serial interface depicted but in fact there might be numerous serial interfaces such as USB, USART or I^2C.

Hardware Component	Description
Sensors	Converts physiological signals to electronic signals. May be on-board the sensor node or attached.
CPU	Responsible for managing system components and data.
Battery	Provides electric power.
Power Management	Manages distribution of power and battery charging.
Data Storage	Stores sensor data in memory such as RAM or flash.
Wireless Tx/Rx	Can upload sensor data to either an online analysis node or offline analysis node when components are available.
Wired Tx/Rx	Can upload sensor data to offline analysis node when components are available.

Table 2.3: Typical sensor node system components

2.3.2 Online Data Analysis Node

In many applications of wireless health sensing systems online data analysis is required. Simply put, online data analysis can provide functionality to monitor the subject's health in

real-time and detect life-threatening emergencies. Consequently, the node must have reliable hardware and software to allow constant access to sensor data, moderate processing power for data analysis and communication to remote servers. The advantage of Android is that its open source application programming interface (API) provides a free software development kit (SDK) for any developer wishing to create an Android application. This allows greater insight into the online data analysis node architecture rather than a closed system.

Since the device is sold with reliable hardware and software pre-installed the wireless sensing system developer can focus the online data analysis node development efforts on the design of sophisticated applications such as:

- Wirelessly pair with sensor nodes via an interface such as Bluetooth.
- Receive sensor data and process it for real-time display on the device screen.
- Leverage additional hardware devices such as GPS to enhance data analysis features.
- Use a wireless interface such as 3G/4G or 802.11 to automatically upload data to a remote server. This can have a large effect on saving device storage space on both the sensor node and online data analysis node.

The Smartphone will have improved performance over an individual sensor node allowing for more powerful program design. Table 2.4 presents the main technical specifications of the recently released Motorola Droid Bionic phone for Verizon Wireless to illustrate this point.

Feature	Capabilities
CPU Speed	1 GHz (dual-core)
RAM	1 GB
Storage	16 GB flash + 16 GB microSD card
Battery Capacity	1735 mAh
Operating Time	650 minutes talk time, 200 hrs standby
Wireless	CDMA (3G), LTE (4G), 802.11b/g/n, Bluetooth 2.1 +EDR
Display	4.3-inches
Additional Features	GPS, E-compass

Table 2.4: Motorola Droid Bionic specifications [19]

2.3.3 Offline Data Analysis Network

The offline data processing network provides access to much greater computing resources, notably computing power and data storage, than the sensor or online data analysis node can offer. Similar to the online data analysis node the design and development effort of this network focuses on using pre-existing hardware and operating systems to build applications that support complex data analysis. The offline data analysis network has several features worth mentioning:

- Transfer of data to this network can be wired or wireless.
- If sensor data is stored on the sensor node it can be uploaded directly to the network, bypassing the online data analysis node.
- The network shown consists of a personal laptop connected to a remote server via an Ethernet connection. Data analysis may occur as either a local software application on the laptop or a web interface hosted by the remote server.

The sharing of information by a web hosted application may benefit both user and developer. A user is not be limited by individual system performance but instead has access to immense computing resources offered by cloud computing. The developer might

also choose to learn more about the sensing system usage characteristics from observation of user statistics. This in turn can guide further system development. Software updates to either the sensor or online data analysis nodes may be distributed by the server. The result is a dynamically evolving wireless sensing system that receives updates to increasingly meet the needs of its customer base.

2.4 Prototype Running Performance System: MDAWN Datalogger

A prototype wireless sensor system, the Medical Daily Activity Wireless Network (MDAWN) Datalogger, was developed to measure running performance in this study. The MDAWN Datalogger system is composed of two accelerometer sensing devices that are worn on the foot and record accelerometer running data. A PC system that captures video of a test subject running is incorporated for system test and validation. The focus in this section is the accelerometer device.

2.4.1 Specification

The MDAWN Datalogger is a tri-axial accelerometer sensing device with a processing unit, onboard flash storage, an analog tri-axial accelerometer and a rechargeable battery. The package measures approximately 1.75 inches wide, 2.75 inches long and 0.5 inches thick. It weighs approximately 0.4 ounces while a typical pair of running shoes weighs around 8 ounces. This makes it suitable to be worn on top of a pair of running shoes. Figure 2.4 shows the MDAWN Datalogger package with the tri-axial accelerometer axes labeled.



Figure 2.4: MDAWN Data-logger prototype

The central processing unit of the device is an NXP LPC2148 microcontroller based on the ARM7 architecture. The main specifications of the microcontroller are given in Table 2.5.

Feature	Capabilities
CPU Frequency	60 MHz
Instruction Width	32 bits
RAM	40 kB
Flash Storage	512 kB
Power Source Voltage	3.7V

Table 2.5: LPC2148 main specifications [20]

The microcontroller is contained on the Logomatic v2 PCB board designed by Sparkfun Electronics as shown in Figure 2.5.

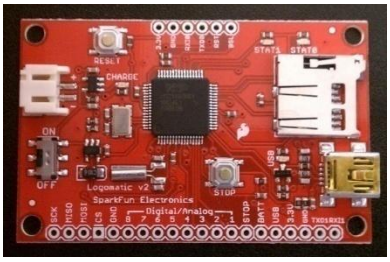


Figure 2.5: Sparkfun Electronics Logomatic v2

The Logomatic board provides access to the following hardware features [21]:

- Built-in real-time clock (RTC)
- 3.3V voltage regulator

- micro-SD socket
- USB 2.0 connection via USB Mini-B
- 2 UART channels
- ADC channels

Open-source firmware that enables data logging capabilities for sensor data collection is available for download. The data logging is user-programmable to select either one of the UART interfaces or the ADC channels. The UART has a configurable baud rate while the ADC channels have a variable sampling frequency. Logged data is stored on a micro-SD flash memory card that is capable of addressing up to 2 GB of memory using FAT16 file system architecture.

The MDAWN Datalogger is powered by an 850 mAh polymer lithium ion battery rechargeable via the USB interface. At the supply voltage of 3.7V lab testing has found that the entire system consumes 70 mA for a theoretical lifetime of approximately 12.1 hours. Consequently it can be expected that the operating lifetime of the sensing node is sufficient for running performance applications.

The accelerometer used in this application is the Analog Devices ADXL326 tri-axial accelerometer. The figures of merit and other sensor features are described in the sensing principles review of accelerometers and provided in Table 2.2. The firmware provides a simple configuration file interface to select the sampling mode (ADC), enable/disable specific ADC channels and set a sampling frequency. In the Datalogger prototype ADC channels 1-3 are enabled corresponding to x, y and z-axis accelerometer measurements respectively. The sampling frequency is programmed to 160 Hz.

2.4.2 Data Management

Sensor data is written out to an SD card on-board the Datalogger. The write time of 42.5 ms makes the data writing the limiting factor in data sampling. With 3 ADC channels active the maximum sampling frequency is 500 Hz [21] so the selected sampling frequency is well within a safe range. Data is written in a comma-separated value (CSV) format such that each row of data contains a timestamp for the sample, x-axis, y-axis and z-axis ADC values.



Figure 2.6: MDAWN Data uploader software interface

After data has been collected by the MDAWN Datalogger the device can be connected to a PC via the USB interface to upload the data. This will also charge the MDAWN Datalogger battery. The Datalogger appears as a removable storage device so data files can be placed directly on the host PC. Alternatively, a software interface has been implemented to upload the sensor data from an MDAWN Datalogger to a central server for analysis. The uploader software interface appears as shown in Figure 2.6. The uploader software has buttons to perform the following tasks:

- *Discover* – discover MDAWN devices attached to the host PC.
- *Upload* – upload sensor data files to the remote server. This also removes the files from the sensing node.
- *Sync* – sync the clocks on the Datalogger devices to the host PC clock for file timestamps.

2.4.3 Sensor Placement

The MDAWN Datalogger has been lab tested to record treadmill and track running data and deployed in the field for use by the Mammoth Track Club (MTC). It has also been now used by six world class athletes in marathons (LA Rock & Roll half marathon and the Honolulu marathon). MDAWN Datalogger devices are placed in running wallet pouches that attach to the top of the running shoes on the laces as shown in Figure 2.7. The choice for the location of sensor was driven by:

- Type of data: Different sensor locations will give different signal values. It is important to identify a location which will provide high amount of quality information. Different locations were tested in practice and it was agreed that top of the foot is a viable location
- Runner comfort: Long distance runners want to run with minimal weight and hindrance to their movement. Upon a small survey, it was decided that sensors attached to the shoes would provide the least distraction or discomfort to the runners



Figure 2.7: MDAWN sensor placement

2.4.4 Sensing System Orientations

1. X-Axis
    - Measures acceleration with positive acceleration in the direction towards front of shoe
2. Y-Axis
    - Measures acceleration with positive acceleration in the left direction
    - Left and right directions are inverted for left shoe processing to produce an acceleration signal in the adduction direction (towards body centerline).
3. Z-Axis
    - Measures acceleration into plane of shoe



Figure 2.8: Sensor system orientation

3 PERFORMANCE METRICS

As discussed in the previous sections, performance evaluation of runners is still an open problem. An elite runner's running technique is shaped by a number of physical characteristics, the influence of previous training (volume, methods of training) and racing conditions. Hence, the running style of each individual is very specific. In order to evaluate the efficiency of a given running technique, we collect data from the 3–axis accelerometer located at the top of the shoe. Performance is then quantized in terms of a few metrics which are derived from this accelerometer data. These metrics give an insight into the flaws and benefits of the existing running technique of a runner and help them better understand their gait cycle.

3.1 Metric characteristics

While every metric must provide distinct and preferably non-overlapping information, there are some features that are desirable in any given metric. Some of these features are:

1. Monotonicity: Perhaps the most important and desirable characteristic of a metric is its monotonicity. The metric should increase (or decrease) with increase (or decrease) in the extent of the underlying physical phenomenon. Please note that monotonicity does not imply linearity. A metric can very well be non linear yet monotonic.
2. Robustness: Ability to handle and extract information from given data in various situations is important. Metric computations should provide monotonic values across different runners under different conditions (like terrain and speed).

Therefore it's imperative that metric values not depend on absolute values of the accelerometer data

3. Low Sensitivity to randomness: Given the nature of the device, inherent randomness in the data is a given. It is essential to have check points within the system to test the sanity of the output values. Use of statistics has proven to be highly helpful in this regard
4. Normalized/Global Values: If possible, metrics should be normalized so that the values hold across all runners. But this might not always be possible given the nature of data and limited degrees of freedom. Efforts are continuously being made to make as many metrics global as possible
5. Low Computational Complexity: Given the fact that almost all metric computations are to be carried out in real time on an Android platform with limited computational power, care has been taken so that metric computations are kept simple
6. Meaningful and Crude: Metrics should convey some useful information to the runner so that the person can make adjustments to the running technique to improve. Some physical interpretation is necessary for the same. At the same time these should be crude enough so as to satisfy all of the above criterion

It is important to mention here that the metrics being devised and explained here, though dependent on the accelerometer data, do not represent actual physical world measurements. Owing to the fact that the accelerometers are attached to the foot and foot of the runner is in constant motion, the reference frame of these accelerometers is not fixed. As a result, the gravity makes it very hard to measure net acceleration of the foot in

two out of the three directions. Hence, it is imperative to strategically formulate metrics that represent the motion in question monotonically but is more or less independent of factors like gravity.

3.2 Metric description

Selection of appropriate metrics is conditioned on metrics fulfilling the criterions mentioned in the previous section. Moreover, these metrics should have an intuitive appeal to them so that runners can interpret this information easily. Hence it is necessary that origin of these metrics be from a physical motion involved in the mechanics of running. A typical human gate cycle can be depicted as shown in Figure 3.1.

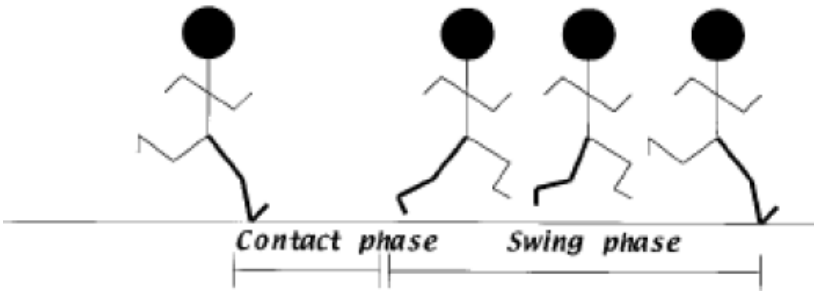


Figure 3.1: Typical human running gate cycle

We can further subdivide the swing phase (as in Figure 3.1) into distal leg lift phase followed by the forward swing phase. We will refer to the contact phase as ground contact phase in this thesis.

Based on the physical features of a typical gate cycle and some common running technique faults, the following metrics have been chosen to represent the running performance:

1. Cadence

2. Ground Contact Time (GCT)
3. Distal Leg Lift measure (DLL)
4. Over-stride measure
5. Contact De-acceleration energy loss measure (CDEL)
6. Degree of Pronation

We will now take a closer look into the meaning and formulation of each of these metrics.

3.2.1 Cadence

Cadence in sports involving running is the total number of 'revolutions per minute' (RPM), or number of full cycles taken within a minute, by the pair of feet, and is used as a measure of athletic performance. The step frequency is often used to model the expended energy in human running using a simple mass-spring model [22], [23]. It is thus an elementary factor describing running kinematics. Elite runner cadence values approximately lie close to 90 steps per minute irrespective of the height of the runners [4].

We use the stride recognition algorithm, discussed in detail in Section 4, to keep track of the stride count and with the system clock at our disposal we can easily calculate the cadence. Here is the mathematical formulation for the same:

Cadence for a given time interval (t) = (Stride Count for time interval t)/t

Here 't' is measured in minutes. We calculate cadence for both left and right foot separately and then average it to obtain the final value for this metric.

3.2.2 Ground Contact Time (GCT)

Perhaps one of the most important, the most robust and the most reliable metric is the Ground Contact Time (GCT). Referring back to Figure 3.1, the contact phase is the ground contact phase of the gait cycle and its timing is called GCT. Thus, GCT is the time from

initial foot strike to toe off. Using the accelerometer data we can calculate the time for which the foot is in touch with the ground. GCT can be presented in two ways:

4. In terms of the fraction of the entire stride – We will call this GCT_F
5. In seconds – We will call this GCT_S

Both of these representations have their own specific advantages and will be mentioned during the course of this description.

There are many studies documenting the importance of GCT. But many of these studies either focus on the variation of GCT with speed or on documenting ground contact times of elite runners. Specifically, the amount of time the runner spends on the ground decreases [24, 25, 26] with speed. Faster athletes are able to apply greater ground forces leading to a shorter ground contact time and greater maximum running speeds [24]. While it is known that as running speed increases, ground contact time decreases, there may also be factors other than speed increases that relate to ground contact time. We confirm the well established relation between GCT and speed, compare GCT's of elite runners but our most important contribution (apart from being able to calculate GCT in real time in race conditions) is leveraging this information to advance our insight into the gait cycle of the runner and pin-pointing the exact features of the running technique.

There are three kinds of runners based on which portion of the foot first touches the ground while landing: fore-foot strikers (FFS), mid-foot strikers (MFS) and rear-foot strikers (RFS). While short distance runners or sprinters are generally fore-foot strikers, long distance runners resort to either mid-foot striking or rear-foot striking. Using GCT metric we can observe the specific trends that each type of running causes and this is what

makes this metric is so fascinating to the running community. Figure 3.2 shows the typical steps a RFS goes through during the ground contact phase.



Figure 3.2: Ground Contact phase of a rear foot striker (RFS)

The initial heel strike is followed by the position where the pressure is evenly distributed on the entire surface of the foot referred to as mid-stance in Figure 3.2. This leads to toe-off which provides the necessary forward push to the body. Studies have documented [27] the adverse impacts of heel striking on the runner's body. The whole body weight plus the running momentum force is applied to the heel in a very short period of time. This impulse force makes the runner more prone to injury. Runners who land on mid-foot are less prone as the same impulsive force is spread across the entire surface of the foot.

In order to develop an algorithm capable of capturing GCT values, we must first examine the accelerometer data. Experiments were conducted and running data was acquired. Figure 3.3 depicts typical z-axis running data between two consecutive foot strikes. The two regions marked by impulses (in red circles) are the beginning of the GCT period. Using this impulsive nature we can easily mark the start of the GCT. This is implemented by leveraging the stride recognition algorithm discussed in detail in the next Section i.e. Section 4.

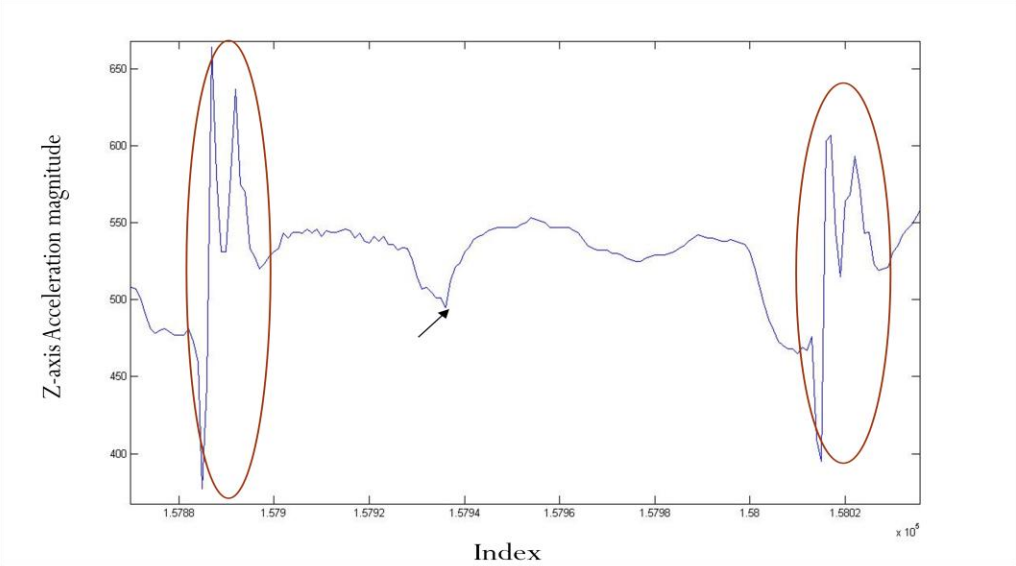


Figure 3.3: Typical z-axis data

Next we need to determine as to which feature marks toe-off. To ascertain this we first need to determine the GCT using other techniques. We use video analysis here of the running session for which the data is being analyzed [13]. The GCT values from video analysis were mapped back to the accelerometer data and it was observed that the local minimum in the z-axis data (pointed at by the black arrow in Figure 3.3) accurately marks the toe-off event in the gait cycle.

Thus, the GCT metric is the timing of the region bounded by an impulse on one end and a negative local minimum at the other end.

As briefly mentioned before, GCT metric can be represented in two ways, namely GCT_S and GCT_F. GCT_S gives an absolute measure of the metric. It helps to compare our results with widely accepted values for GCT and thus acts as a good feedback. Lower GCT_S values point to a higher efficiency in the gait cycle and therefore this value is essential for a runner to track his/her performance and also compare it to elite runners. While this in

itself is complete information, GCT_F also plays a unique role in providing additional insight into a runner's gait cycle. Again, the value of GCT is generally affected by abnormalities in the running technique. And GCT_F is particularly effective here. An unusually low or high value of GCT_F signals a source of inefficiency in the biomechanical technique; a normal GCT_F value may not necessarily imply perfect technique. Studies have documented that lower values of GCT_F mean higher running performance if comparison between runners is conducted [3], [28]. Therefore, GCT can act as a pre classifier to many important flaws. To highlight this point, we leverage the different types of running flaws. We introduced varying degrees of technical flaws in a normal running stride one by one and recorded the data. Upon analysis, the GCT_F values revealed an interesting trend. Average values for the same are recorded in Table 3.1. Please note that these value ranges do not belong to an elite runner; experiments were performed only to determine the monotonicity of the metric.

S. No	**Type of Run**	**GCT_F**
1.	Major Under-stride/Major Low leg lift	0.38
2.	Minor Under-stride/Minor Low leg lift	0.364
3.	Normal	0.3598
4.	Minor Over-stride/Minor High leg lift	0.321
5.	Major Over-stride/Major High leg lift	0.268

Table 3.1: GCT_F values for runs with different technical flaws

We can observe that the effect of different types of runs is being captured by the GCT_F metric. The fact that major under stride or major low leg lift run would have a higher GCT_F as compared to a major over stride run is only logical and this is correctly captured by the GCT_F metric. Figure 3.4 below depicts the monotonic nature of GCT_F values for the different types of runs listed in Table 3.1.

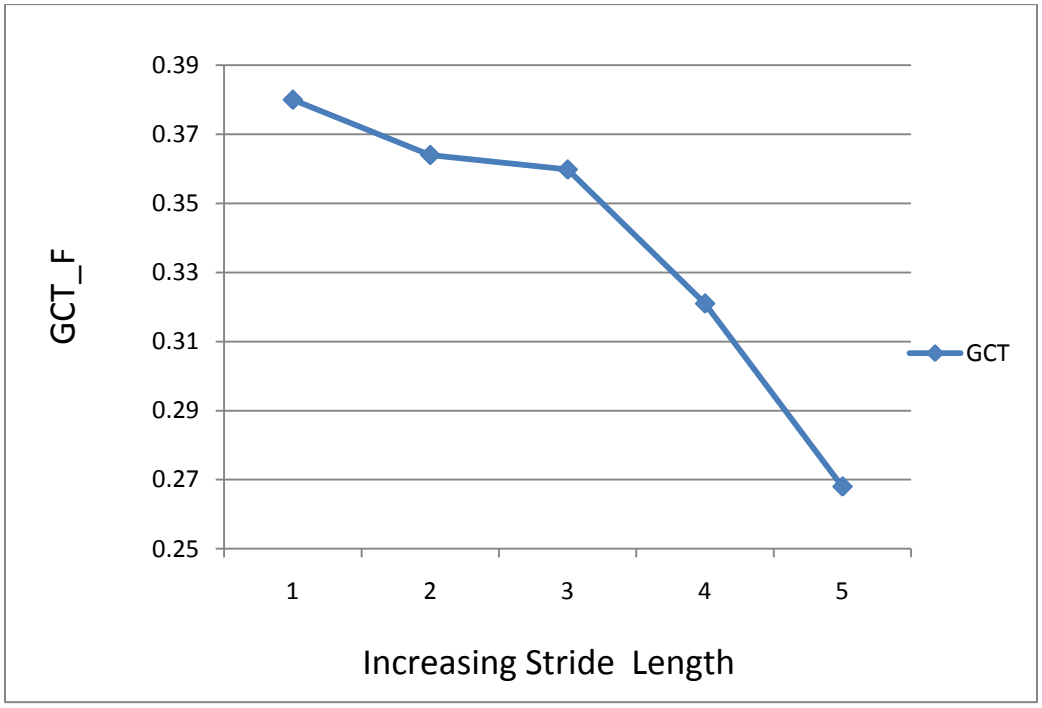


Figure 3.4: Graph highlighting the monotonic nature of average GCT_F values for the 5 cases mentioned in Table 3.1

Please note that this information cannot be captured by GCT_S since it's an absolute measure while GCT_F is affected by rest of the stride as well. Thus, GCT_F can be used as a pre-classifier. A high GCT_F value can be because of any of the below or a combination of these:

1. Flaw in the foot landing technique

2. An unusually high/slow distal leg lift
3. An unusually long/slow forward swing

And low GCT_F values indicate the opposite. Thus knowledge of both GCT_S and GCT_F is important and provide distinct yet relevant information.

This pre-classification coupled with other metric computations can be used to pin point the inaccuracy in the running technique. Now we take a look at some other metrics designed for specific portions of the gait cycle.

3.2.3 Distal Leg Lift (DLL)

Following the ground contact phase is the distal leg lift phase. From toe-off the leg kicks back and flexes. This phase ends when the leg starts coming down again leading into the forward swing motion. It has been well documented that the angle of flexion plays a key role in the running performance. Higher leg lift has been shown to lead to a higher running efficiency [4], [29].

Much research done previously used flex sensors but many have used advanced 3D motion analysis techniques to capture the extent of leg bent or the Distal Leg Lift (DLL). It was our endeavor to represent DLL using a metric that would be increase monotonically with the extent of leg lift using only the available accelerometer data.

Using trial sets of data and careful video analysis, it was observed that there are two things that are play a vital role in determining DLL from the tri-axial accelerometers:

1. Rate of change of acceleration: The accelerometer provides us with instantaneous accelerations. Rate of change of acceleration provides the metric with the intuition of extent of leg lift in a given time period.

2. Time window for DLL: To determine the time for which this phase of gait cycle lasts, we need the toe off point and the point where forward swing just starts. We already have the toe off point from the GCT metric but locating the marker for the end of leg lift phase is a non trivial problem. The accelerometer data seamlessly transitions into the forward swing phase without a distinctive feature to mark the transition.

Given the lack of a distinguishing feature, we have to rely on heuristics to mark the end of DLL period. Much data was gathered and video analyzed, and it was found that for many strides the DLL period ended around 0.78 fraction of the whole stride. This calculation may not be exact but given the constraints and the available information, it was found to be the best solution. Upon implementation it was found that this was a rather fair assumption. Consider for example, a low DLL stride. It can be debated that a low leg lift would result in the DLL phase ending sooner and the 0.78 assumption should fail. But recall that the DLL timing is in fraction of the whole stride – low DLL leads to reduced timing for the entire stride which in turn increases GCT_F. These consequences greatly affect the starting point of the DLL window (due to changes in GCT_F) but have little impact on the finishing point of the window. A similar reasoning holds true for a stride with high leg lift.

Once the window for the metric is fixed, we use the two points above to formulate a monotonic metric. In order to quantize this metric, we use x-axis accelerometer data. Since the available data is acceleration, the acceleration spread is a function of the extent of DLL and the time for which DLL takes place.

$$DLL = (\Sigma^{w}|\text{first derivative of x axis accelerometer data}|)*w$$

Where, w = time window of DLL = (0.78 – GCT_F)*(stride length in seconds)

Please note that we make use of the x-axis data only since this is the axis along which the motion takes place. We ignore the lateral movement of the foot in this metric as it is inconsequential and hence unnecessary. We back tested this metric with the data available and shown below are some basic results highlighting that the metric is indeed monotonic.

	Major Low Leg Lift	**Minor Low Leg Lift**	**Normal Leg Lift**	**Minor High Leg Lift**	**Major High Leg Lift**
Average values - Alana	2872	2978	4190	5893	20245
Average values- Andrew	4618	7062	10835	17823	36662
Average values - Yeung	-	3851	4629	18335	-

Table 3.2: DLL metric values for different types of runs for the 3 runners

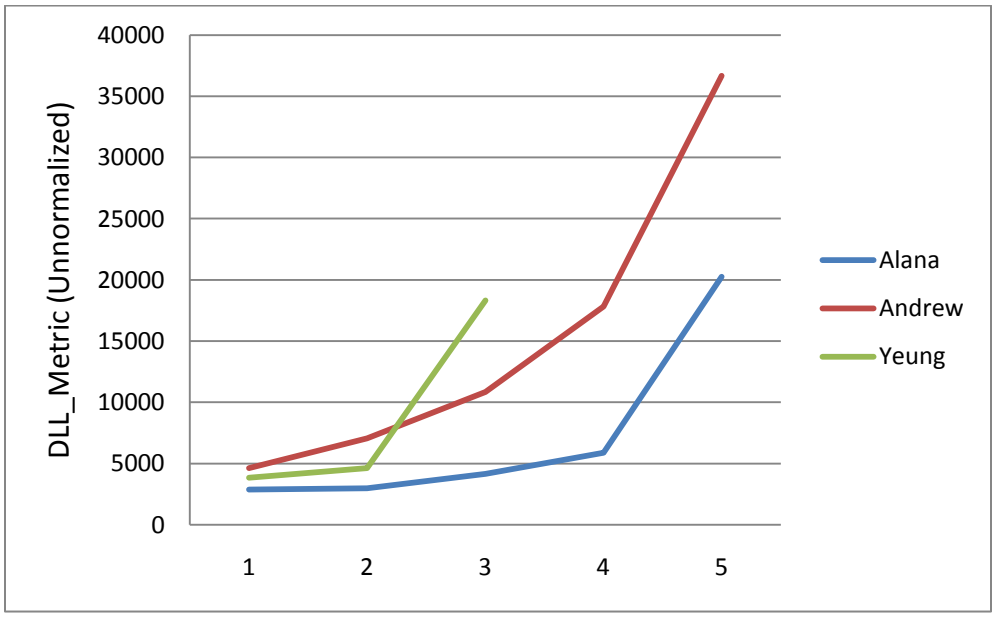


Figure 3.5: DLL metric vs. the increasing degree of leg lift

Clearly, the DLL metric has a monotonic relation with DLL for each dataset. As can be seen from Table 3.2, with increasing degree of leg lift the value of the DLL metric also increases. Figure 3.5 plots this monotonic relation for all the runners. But, the value for each metric is not consistent across different runners. Thus it is not a global metric i.e. for a given extent of DLL, it can potentially give different values. Runner height is one of the reasons for such a difference. Data from several runners can be used to construct a statistical model leading to a normalized DLL metric. This metric has shown very high consistency and is a major contributor to the overall running efficiency metric.

3.2.4 Over-stride Metric

One of the most common injury-causing running form mistakes is over-striding which leads to landing heel first with your foot well ahead of your body's centre of gravity. Some runners assume that a longer stride will improve their speed or running efficiency, but that's not the case. Over-striding wastes energy since it means you're braking with each foot strike and thus stopping the momentum. It could also lead to injuries such as shin splints. It is necessary that runners don't lunge forward with their feet. This is especially important when running downhill. Focus should be on landing mid-sole, with the foot directly underneath the body with every step.

The end of the distal leg lift phase naturally leads into the forward swing. The leg stretches in front of the body. The end of this event is marked by the foot touching the ground. In order to quantize the extent of the forward swing, we use concepts developed while discussing the DLL metric. Like in the DLL metric, we are essentially trying to capture the extent of leg movement but in this case in the forward direction. It is important to mention here that this metric gives a continuous output and not a binary decision.

Output is monotonic with the extent of the forward swing. This metric has similar normalization needs as the DLL metric. The lower cut-off for the time window of this metric is the end of the DLL metric time window which, as you may recall, was a heuristic with a value of 0.78*(stride length). This section ends when the foot is just about to touch the ground again. Hence,

Over-stride Metric = (Σ^{w}|first derivative of x axis accelerometer data|)*w

Where, w = time window of forward swing = (1 – 0.78)*(stride length in seconds)

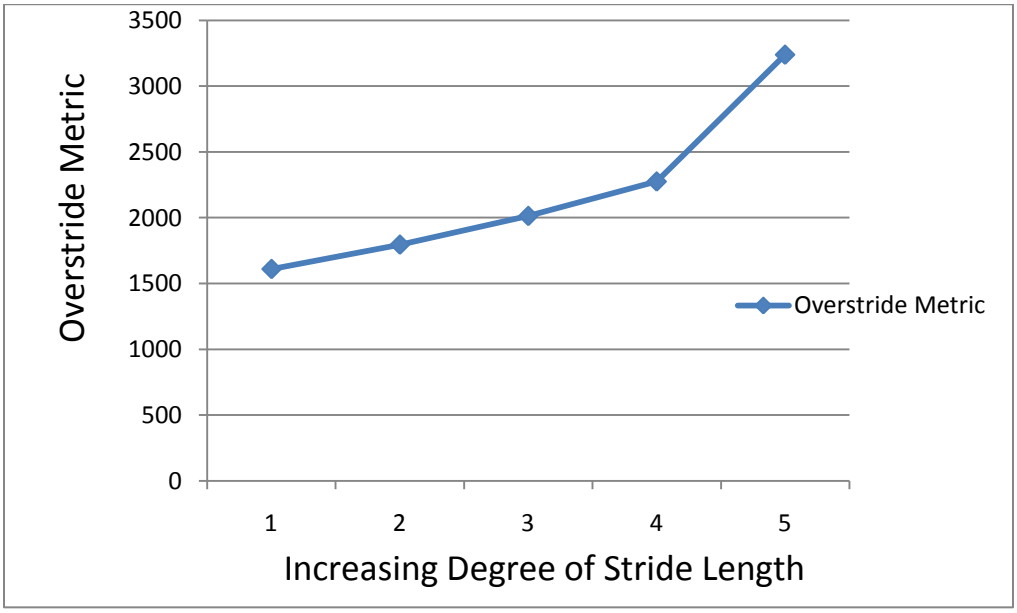


Figure 3.6: Over-stride metric vs. increasing degree of stride length

Experimental runs were performed with increasing degree of stride length. Average values for each run are plotted in Figure 3.6. Clearly, the metric is monotonic in nature.

3.2.5 Contact De-acceleration Energy Loss (CDEL) Metric

This metric is a measure of the ground reaction forces acting on the foot when it lands on the ground. Higher reaction forces while landing means higher impulse which in turn put

larger pressure on the body and thus this is definitely uncalled for. Higher ground reaction forces also have a negative impact on the running economy [30]. Thus lower the reaction forces are better for high performance. It is important to monitor these forces. We call this metric Contact De-acceleration Energy Loss (CDEL).

A good runner slows the foot down just before landing in order to reduce the reaction forced and thus reduce CDEL. The region of interest here is the data segment corresponding to and just after the initial ground contact. Also, the direction of interest is the vertical direction, since most of the reaction is in the vertical direction. We therefore focus on the z-axis accelerometer data for this metric development. Measurement of change in acceleration during the initial foot landing period will give us a measure of the disturbance experienced by the foot. This is admittedly noisy data because the accelerometer is itself not perfectly still at the point of impact. But the underlying assumption is that this noise is randomly distributed and effect of specific trends will emerge out of data regardless. Thus,

$$\text{CDEL Metric} = (\Sigma^{t}|\text{first derivative of z axis accelerometer data}|)$$

$$\text{Where, } t = \text{time window of CDEL Metric} = 0.1 \text{ seconds}$$

We have chosen a time window of 0.1 seconds as it was observed from the data that most impulsive forces died out after this period. Please note that this metric is not normalized by any variable value. It has been observed that this metric is affected by the weight of the runner but the relationship is not linear. Data from several runners might help in formulating a statistical model which can result in emergence of a normalizing factor which will make this metric global.

Experiments were then conducted to confirm this hypothesis. A runner was asked to run with varying degrees of 'thump' while landing. We call the force while landing on the ground as 'thump'. Logically, higher thump runs will generate more ground reaction forces and this should lead to a higher CDEL metric for a monotonic metric. On inspection, this was found out to be true. Table 3.3 gives the metric values for 5 Normal runs and 5 heavy footed runs. CDEL values for the heavy footed runs were found to be more than normal run values.

S. No.	**CDEL for Normal Run**	**CDEL - Heavy Footed Run**
1	423	648
2	489	868
3	470	711
4	463	748
5	512	687
Average	471.4	732.4

Table 3.3: Average CDEL metric values for 5 Normal and 5 Heavy footed runs

3.2.6 Pronation

The landing is one of the most important parts of the stride. Large magnitudes of impulsive forces are experienced by the runner when the foot touches the ground. Thus it is important to make contact in a way which minimizes these forces. In a normal stride, the

outside part of the arch of the foot makes initial contact with the ground. The foot "rolls" inward about fifteen percent, comes in complete contact with the ground, and can support your body weight without any problem. The rolling in of the foot optimally distributes the forces of impact. This movement is called "pronation," and it's critical to proper shock absorption. At the end of the gait cycle, you push off evenly from the front of the foot.

In proper terms, pronation is the movement of the sub-talar joint (between the talus and calcaneus) into:

- Eversion (turning the sole outwards)
- Dorsiflexion (pointing the toes upwards) and
- Abduction (pointing the toes out to the side)

The most common defects in running technique related to pronation are over pronation and under pronation. As the names suggest, these defects are caused by over rolling or under rolling the foot respectively just before contact. Excess stress on the inner surface of the foot through over pronation can cause injury and pain in the foot and ankle. Repeated rotational forces through the shin, knee, thigh and pelvis also place additional strain on the muscles, tendons and ligaments of the lower leg.

Now in order to mathematically represent this metric, we either need a direct measure of the rotation of the foot or a representative metric of the same. Using the y-axis accelerometer data, we try to develop this metric as the rotation data is best captured in the y-direction. A double integrated time series of the y-axis accelerometer data gives us a distance vs. time graph representation (Please note that we cannot calculate exactly the distance traveled using only this data). And the rotation of the foot prior to landing is proportional to the distance traveled in the y-direction. Thus we have important

information about the extent of pronation of runners from these graphs. To bring about a change in the degree of pronation is a slow process and therefore this metric was suitable to being kept offline. Hence the graphical nature of the metric is justified and we can analyze the graphs at a later time.

In order to validate the theory, data from a tempo run of Josh Cox (an elite runner) was used. Video analysis techniques [13] were used to confirm the conclusions drawn from the results below. Integration was done for the duration of the ground contact time of the gait cycle. Figure 3.7 shows y-axis abduction data from the accelerometer. It is worth noting that y-axis data has much lesser amplitude swings than the other axis since most of the acceleration takes place in the x or the z-axis over the course of a normal gait cycle.

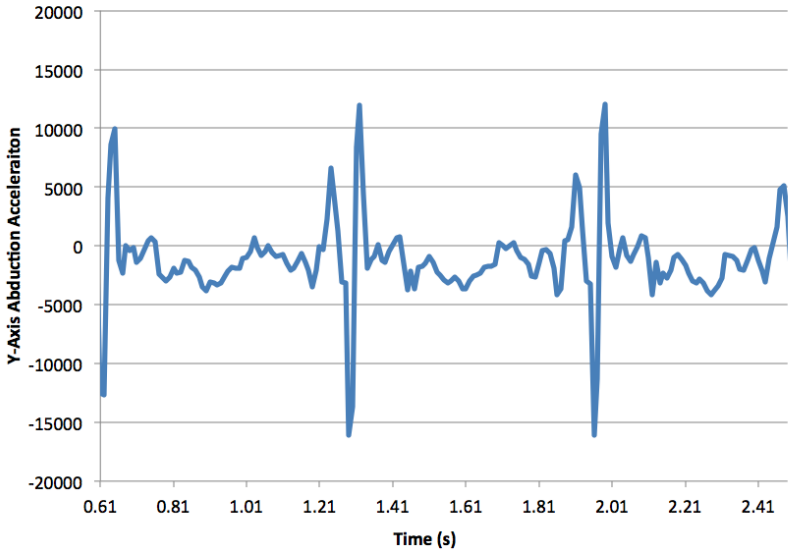


Figure 3.7: System Acceleration in Abduction Direction for Left Shoe in Tempo Run

Figure 3.8 depicts pronation rotation metric vs. time graphs for the left and right foot at different points in the tempo run. Clearly, the rotation of the two feet is in different direction and there is a remarkable difference in the degree of rotation. As can be seen, the

data is not symmetric. We can see effects of pronation and supnation in the left foot data (confirmed by video analysis) and only pronation for the right foot. These results are interesting for a runner to look at in order to understand his technique better.

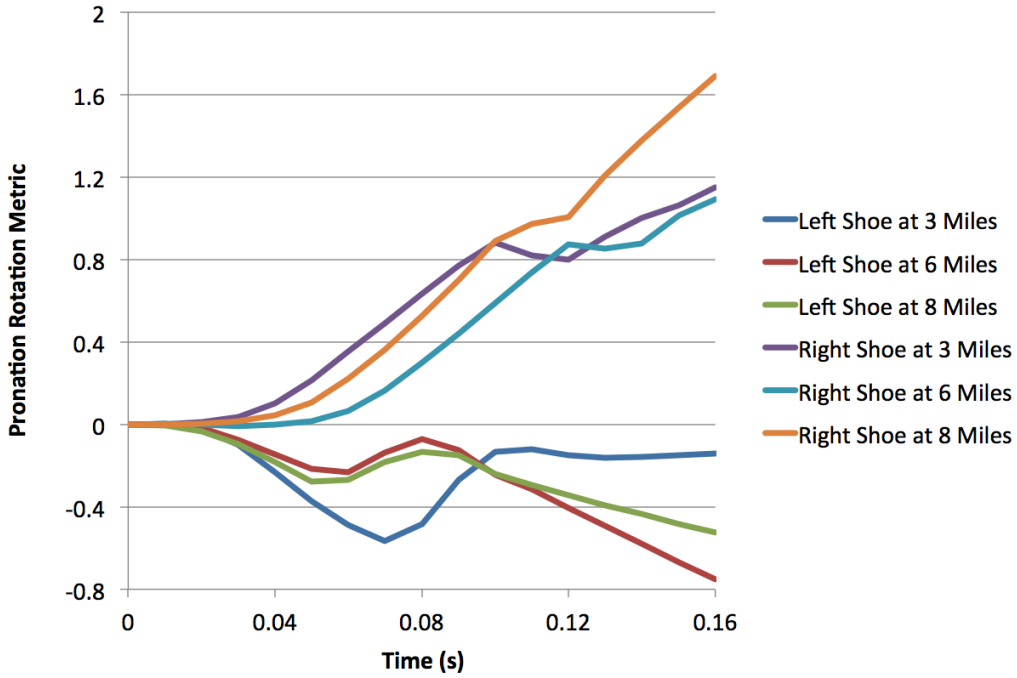


Figure 3.8: Combined Left and Right Foot Pronation Rotation at different stages of the run

4 ALGORITHM ARCHITECTURE AND STRIDE SEGMENTATION

To effectively design an efficient algorithm it is important to keep certain basic steps in mind. An efficient real time algorithm should be able to handle and eliminate false positives, keep false negatives to a minimum, be computationally light and must present the output in a meaningful manner. Keeping these constraints in mind, a basic work flow was designed for development of the algorithm architecture. Refer to Figure 4.1 for the high level flow chart of the same.

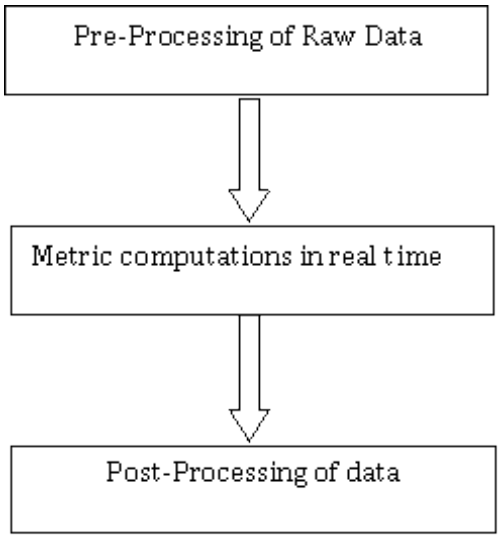


Figure 4.1: Work flow for algorithm design

Any raw data being captured must first be processed and checked for false positives and it should be ensured that the values being recorded are within a certain expected range. This is followed by specific metric computation algorithms as discussed in the previous section. Finally, a sanity check is necessary before presenting the output values to the user. Now let us examine each of three steps shown in Figure 4.1 in a little detail.

4.1 Pre-processing of raw data

The raw data must undergo some processing before we can apply metric algorithms on it. The sensors record the acceleration and these values are stored. Please recall that all the metric formulations described in Section 3 are applied on data segments of one stride each. Hence as a first step it is imperative that we divide the raw data chunk into segments of one stride each. This procedure is called stride segmentation and is the first step in our implementation. This is then followed by isolating running data and discarding data resulting from performing any other activities.

4.1.1 Stride segmentation/recognition

This is essentially the first step in algorithm design. The raw data being recorded on the fly is being first passed through this segmentation exercise. The design of this algorithm is essential for everything else that follows. The raw data looks something like as shown in Figure 4.2. The figure shows acceleration vs. time index plot of a trial run.

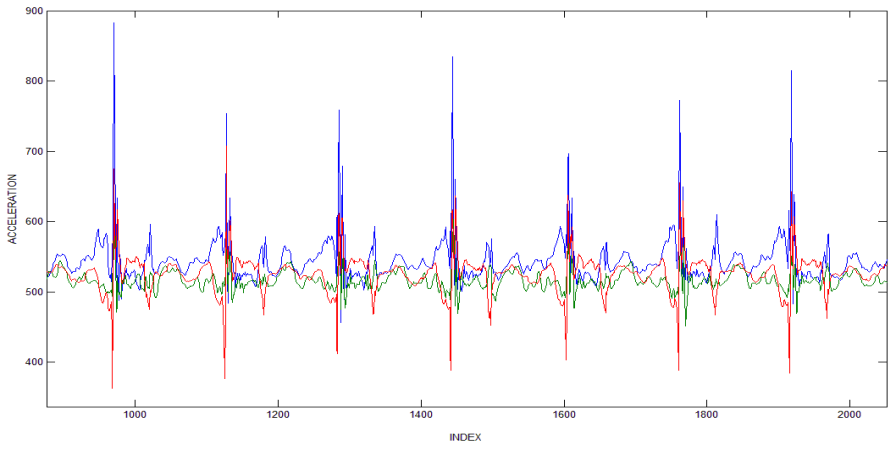


Figure 4.2: Raw acceleration data from the tri-axial accelerometer

As can be observed, this continuous chunk of data has many strides. In order to separate each stride we have to segment the data based on a feature. Observe the areas

characterized by spikes. These are spikes caused by the rapid de-acceleration resulting from the foot striking the ground. This feature has been exploited to segment the data. Specifically, we calculate the first derivative of the raw data being recorded and sum the derivative within a moving window of pre-specified length. This window is glided over the entire length of the data. Figure 4.3 displays how the resulting data vector looks like. A window of length three was chosen for the purposes of stride recognition and is kept fixed at all times. This value is based on trial and error methodology applied on training data.

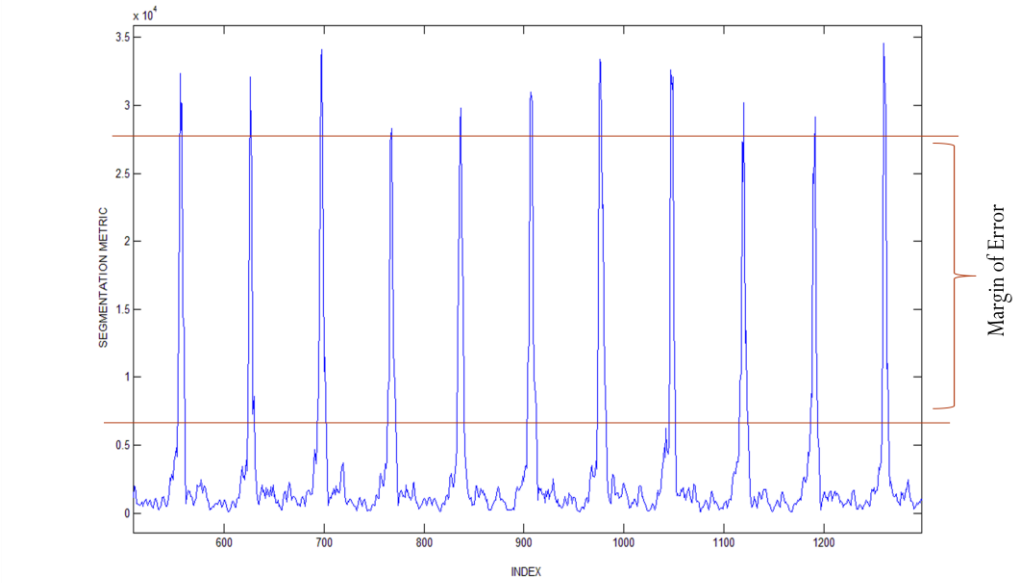


Figure 4.3: Data vector generated for stride segmentation purposes highlighting the margin of error

The resulting outcome is then the magnification of the effect of rapid de-acceleration caused by foot landing on the ground. Due to this magnification, we can now easily mark the points in close vicinity to the initial stages of ground contact by using a threshold value. This approach has two advantages:

1. Any random spike in the middle of the data series would not lead to a false positive as the summing over the sliding window averages over the effect of any such event.
2. The difference in the amplitude, as is clear from the figure above, is exaggerated thus providing a higher margin for error. It was observed, especially for female runners and light weight runners that the foot landing event is not as pronounced as in other runners. But this approach exaggerates the foot landing event on account of the fact that the data window near that region has a very high first derivative.

While the above approach is very effective, it might not be fool proof. Hence additional measures must be taken so that proper segmentation is done and false positives are kept to a minimum. One such technique is based on the idea of exploiting expected cadence values. After a foot strike is detected, based on the widely accepted cadence range for running, a portion of the data is not passed via the segmentation algorithm. This then prevents any other event during the stride from interfering with the segmentation algorithm. This was particularly useful for runners with 'soft landing' or in other words runners with a less pronounced foot landing event.

Referring back to Figure 4.3, another important part of the process is deciding the value of the threshold. The threshold value should have the property of being consistent and constant for all runners. The margin of error created by taking the first differential and summing the result over a window resulted in the evolution of a global threshold value for all runners. This value can be upwards of 15000 (19000 was used in the actual algorithm) for the acceleration values recorded by the MDAWN sensors.

Stride recognition, in summary, exploits the high first derivative of the initial ground contact region, threshold detection and time-gating based on expected cadence.

4.1.2 Activity Detection

It is important to make sure that the data being processed is indeed running data. Any runner/athlete might do activities other than running while wearing the sensors like jumping, walking, stretching etc. It would be unnecessary and a waste of resources (for example battery life and computational power) to run metric computations on this data. Hence, there needs to be a check to differentiate between running data and all other data. It is worth mentioning here that this can also act as an activity detection marker.

To segregate running data, we leverage the segmentation algorithm designed above. Please recall that the segmentation algorithm is designed to isolate data between two consecutive foot strikes. Also, the threshold can be adjusted so that only running steps are detected. This is essence means that the time difference between two consecutive foot strikes while running can be calculated. We can then calculate the cadence based on this time difference. If the cadence is not in the usual and accepted range for running, then that data is discarded. The acceptable cadence range used in our algorithm is 65-120, which is widely accepted. Consider, for example, the activity of walking. Walking data looks very similar to running data pattern but the amplitudes of acceleration are lower. This fact has been used to design the threshold in a way so that only running is detected and walking is not detected. Now consider jumping. It will create data spikes similar in nature and magnitude to the ones generated by the event of foot landing on the ground while running. We do not want to misclassify jumping as running because it will disturb the metric averages being calculated and hence lead to misinformation. So to tackle this problem, we

basically use cadence as a check point. Jumping will lead to a very high cadence which will fall out of the cadence range discussed earlier. Essentially this process dramatically reduces the false positives. This thus helps to isolate only running data and only this data goes through the second stage i.e. metric calculation.

4.2 Real time metric computation

Data of all the three axis of the tri-axial accelerometer, in segments of one stride each, is used for metric calculation. These are metrics as described in section 3.3. Because of limited computational power there were mainly two concerns that needed attention:

1. The choice of frequency of MDAWN sensors: 100Hz or 200 Hz. While 200 Hz provides more resolution but a higher frequency means more data points per second and given the limited computational power, we decided to use 100Hz as the default frequency value.
2. Based on the computational complexity, some metrics can be calculated online while others can be computed offline. While most of the metrics are presently online, degree of pronation is a metric which is kept offline as it is a graph and adds little value in real time.

Metrics of left and right foot are calculated separately and kept independent of each other.

4.3 Post processing of data

The two steps described above feed metric values into this stage of the algorithm. Every metric value must go through a 'sanity check' or in other words, metric values must lie within the accepted range of their respective values. This is where use of statistical tools can prove to be very useful. Based on the standard deviation and mean of each metric

value, we can come to a fairly accurate conclusion of whether the value in question is representative of the metric or not. For example, consider the GCT metric. We know that its value for a good runner should lie between 0.13 s to 0.25s. Any value much beyond this range is suspicious and must be weeded out. A point worth mentioning here is that there can be many sources of error in these metrics. Some of the common error sources are:

- False positives which evaded the weeding process at the pre processing of raw data
- False negatives: Although inherently the algorithm generates very few false negatives (A running stride not being classified as a stride due to an error in the segmentation algorithm), there is still a possibility of the same
- Hardware issues
- Inability of the algorithm to handle a possibly rare data pattern

Once each metric value is checked for 'sanity', we need to present these metrics to the runners in an easily readable format. For this purpose we use a moving average wherein we average the metrics over time and display the instantaneous values as well as long term averages of the metrics. This way a runner can track and compare his performance history throughout the race. Hence the output of the system is the moving average of metric values, each of which provides a unique insight into the biomechanical running technique of the athlete.

5 REAL TIME DEPLOYMENT AND RACE RESULTS

In order to deploy the system in real time, thoughtful and careful deliberation was required to set up the system architecture. While system design decisions pertaining to the metric calculation have been discussed in detail in Section 2, several key questions regarding data storage and real time information display still needed attention.

For continuous monitoring of a runner's performance in real time by coaches and possibly commentator's, we need a system so that raw data as well as metrics calculated were uploaded onto a central server which could then be accesses using internet at any remote location. Here is a brief review of the system set-up:

1. The MDAWN data logger which is a tri-axial accelerometer is placed on the foot of the runner in an almost horizontal position
2. The raw data is stored locally. For storage and battery life purposes, the frequency of the sensor is selected to be 100 Hz
3. No computation is done on the MDAWN platform although there are signal processing capabilities
4. Data is streamed to an Android based Smartphone which the runner carries via Bluetooth technology
5. Metric calculation algorithms (in java script) are put to use on the Smartphone. The input to these algorithms is the raw data and the output being the metrics discussed in Section 3

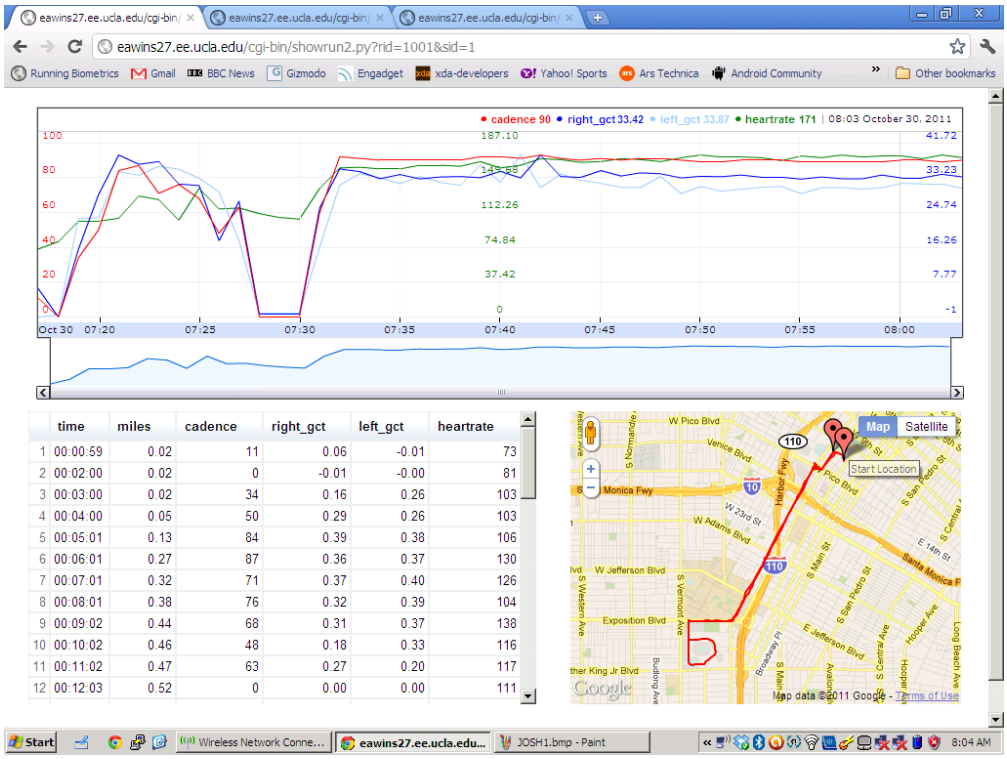
5.1 Real Time Deployment

Now given the existing system, we have two more system requirements that need attention:

1. Logging raw data at a remote location in order to keep the Smartphone memory relatively free
2. For real time runner performance evaluation and comparison, we need to display these metrics on a remote device. We thus need to upload these metrics as well to a remote location

We can here leverage the Smartphone; we set up a system where after a fixed interval of time, the data and the metrics are messaged using the SMS (short message service) capabilities of a Smartphone. The fixed interval was chosen to be 500 seconds for logging raw data and an even smaller interval for metrics.

While there is no need to immediately access the raw data, we need to display and analyze the metric data during the race. To achieve this, a real time display system was built by one of the team members which using metric data from the server, displayed all the key information in a compact and user friendly manner. It is worth mentioning here that we also leveraged the GPS (Global Positioning System) capabilities of the Smartphone in order to track the location of the runner. GPS data, though, was not used for any other purpose. Figure 5.1 displays a screen shot of the real time display system used for the purposes described above during the LA Rock and Roll Half Marathon. It displays metric values, GPS location and metric graphs for each runner.



	time	miles	cadence	right_gct	left_gct	heartrate
1	00:00:59	0.02	11	0.06	-0.01	73
2	00:02:00	0.02	0	-0.01	-0.00	81
3	00:03:00	0.02	34	0.16	0.26	103
4	00:04:00	0.05	50	0.29	0.26	103
5	00:05:01	0.13	84	0.39	0.38	106
6	00:06:01	0.27	87	0.36	0.37	130
7	00:07:01	0.32	71	0.37	0.40	126
8	00:08:01	0.38	76	0.32	0.39	104
9	00:09:02	0.44	68	0.31	0.37	138
10	00:10:02	0.46	48	0.18	0.33	116
11	00:11:02	0.47	63	0.27	0.20	117
12	00:12:03	0.52	0	0.00	0.00	111

Figure 5.1: A screenshot of the real time display

5.2 Race Results

The goal in distance running is to find a balance between maximum sustainable speed and energy production. This balance plays a vital role in distance running. Throughout the years, athletes have sought for ways to improve performance in long distance running events by balancing the distance of the race with speed. Marathons are certainly a very unique opportunity to capture and analyze real time running data of elite runners in a competitive environment. The results presented in section 5.3 and 5.4 describe in brief some of the important findings and how they have advanced our understanding of running biomechanics.

5.3 Los Angeles Rock & Roll Half Marathon

The first real time deployment of the product was carried out at the LA Rock & Roll Marathon. Following are the details:

- Event: Half marathon
- Location: Los Angeles, California, USA
- Distance: 13.1 miles
- Date: 30th October 2011
- Course Elevation Map

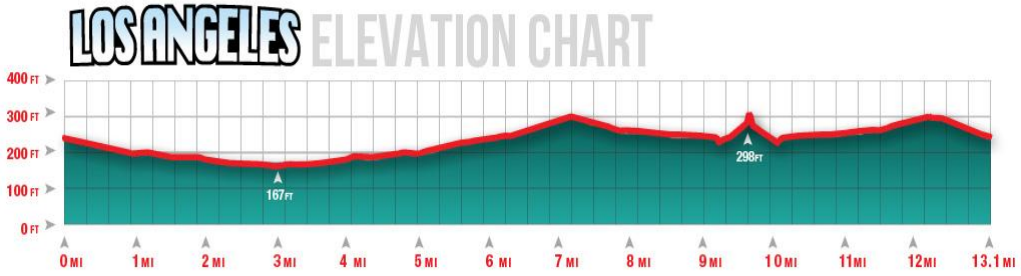


Figure 5.2: Course Elevation chart for the LA R&R Half Marathon

- Participating Runners: Three accomplished runners participated in the program and two of them finished in the top fifteen. Their names are as follows:
    1. Josh Cox
    2. Jonathan Pierce
    3. Brian Sivy

Before we go on to examine each metric result in detail, it is important here to mention the actual race results. We will be referring to these during metric comparisons for the three runners. Please note that the actual results were obtained at the end of the race and the metric results were computed in real time.

5.3.1 Final Race Results

Having analyzed the performance of the runners using accelerometer data and the developed algorithms, it is important to have a feedback system which confirms our findings and analysis. Actual race performance is a very good indicator for the same. Figure 5.3 shows the race timing comparisons for the three runners.

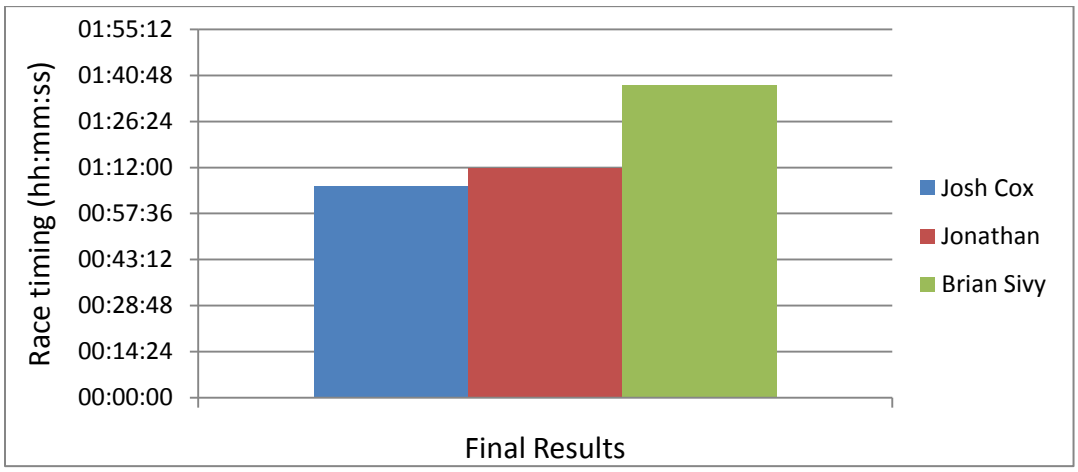


Figure 5.3: Final race timings and comparison of the three runners

As can be seen, Josh Cox performed the best and finished 5th with a timing of 01:06:05. He was followed by Jonathan Pierce with Brian Sivy finishing well behind these two. If our metric comparisons indeed reflect actual race performance, then we shall be able to see these differences in metric values. We shall now look at the results of each runner and their metric performance in detail. Please note that owing to a fault in the data-logger, John's left foot data was recorded for only a portion of the race and hence you will see unequal time series data for his left foot.

5.3.2 Cadence

Let us first examine the cadence of all the three runners over the length of the entire race. Figure 5.4 gives the cadence values for the entire race for all the three runners.

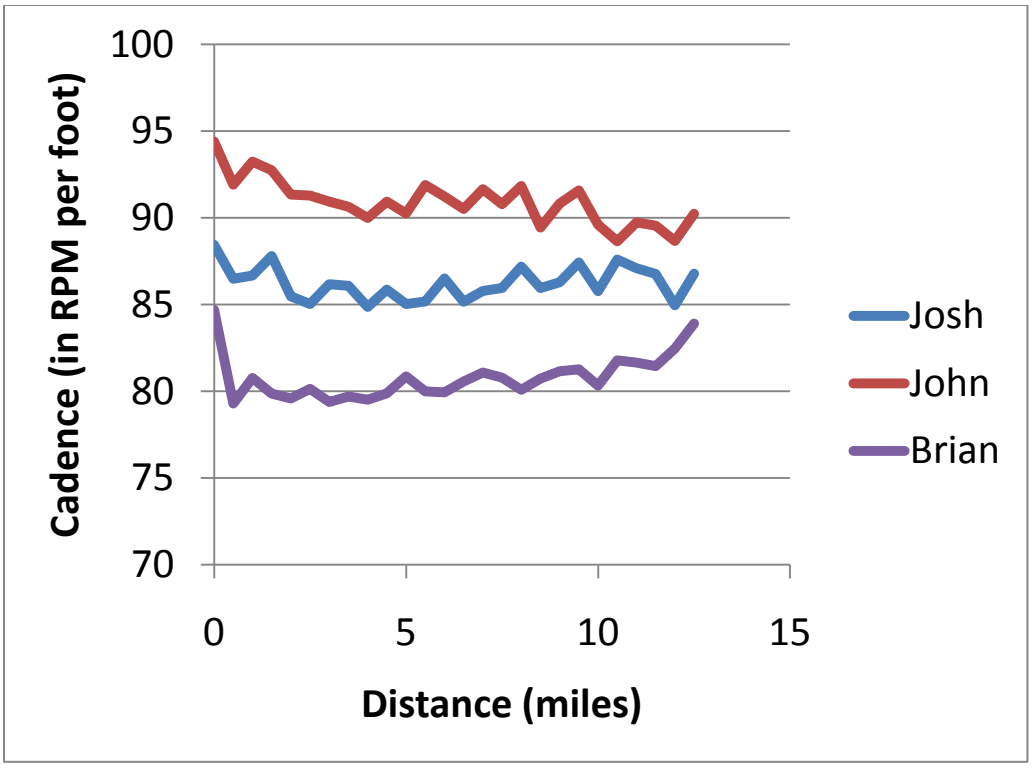


Figure 5.4: Distance vs. Cadence time series for the three runners

A few preliminary observations and conclusions from the graphs above are:

1. Josh and John fall within the range of 85-95 rpm while Brian's cadence is much below that. This is a good indicator of the race performance because as we now know that Josh and John were top performers while Brian lagged behind
2. John's cadence falls off drastically as the race progresses while Josh's cadence remains rather constant. This is also an important result because upon analyzing the race performance over the course of the race it was found out that John's performance indeed dipped while Josh was able to keep up with the race leaders at all times

3. It is also interesting to note that there is a small increase in the cadence of all the runners near the end of the race. This can be attributed to the psychological impact of the finish line

From the observations and conclusions for this metric and the actual race results, we can easily observe that we are, to some extent, able to track the performance using this metric.

5.3.3 Ground Contact Time (GCT)

As discussed in Section 3, GCT (GCT_S and GCT_F) is one of the most important metrics for performance tracking of runners in real time. Refer to Figure 5.5, 5.6, 5.7 and 5.8 to see the time series graph for all the three runners for this metric.

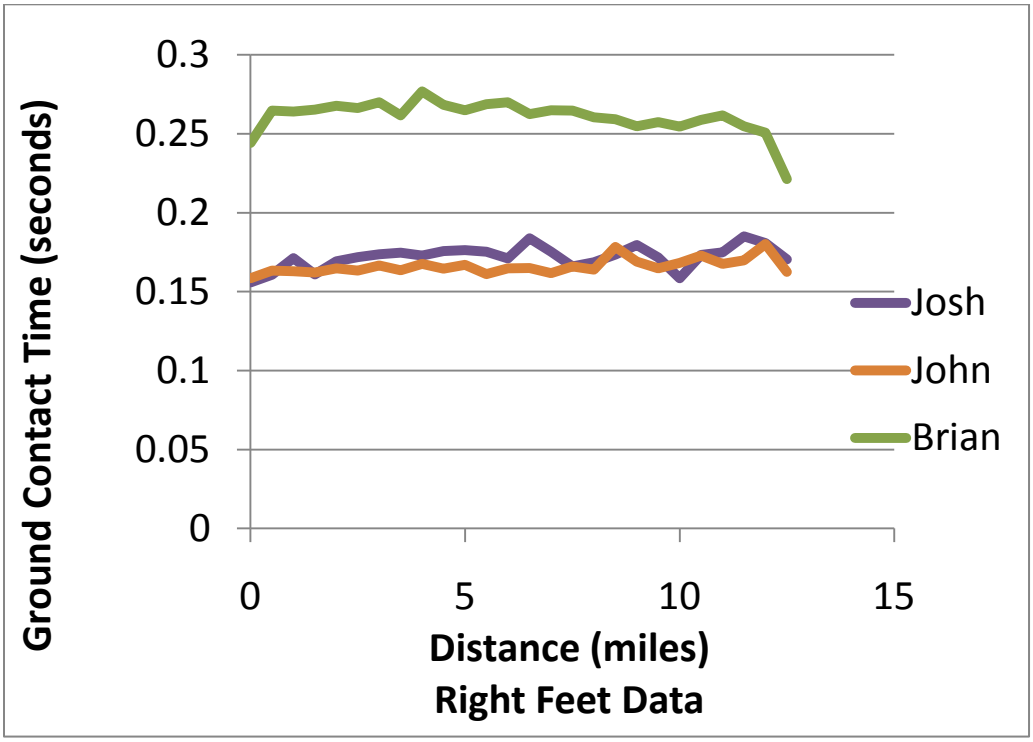


Figure 5.5: GCT_S of right feet vs. Distance for the three runners

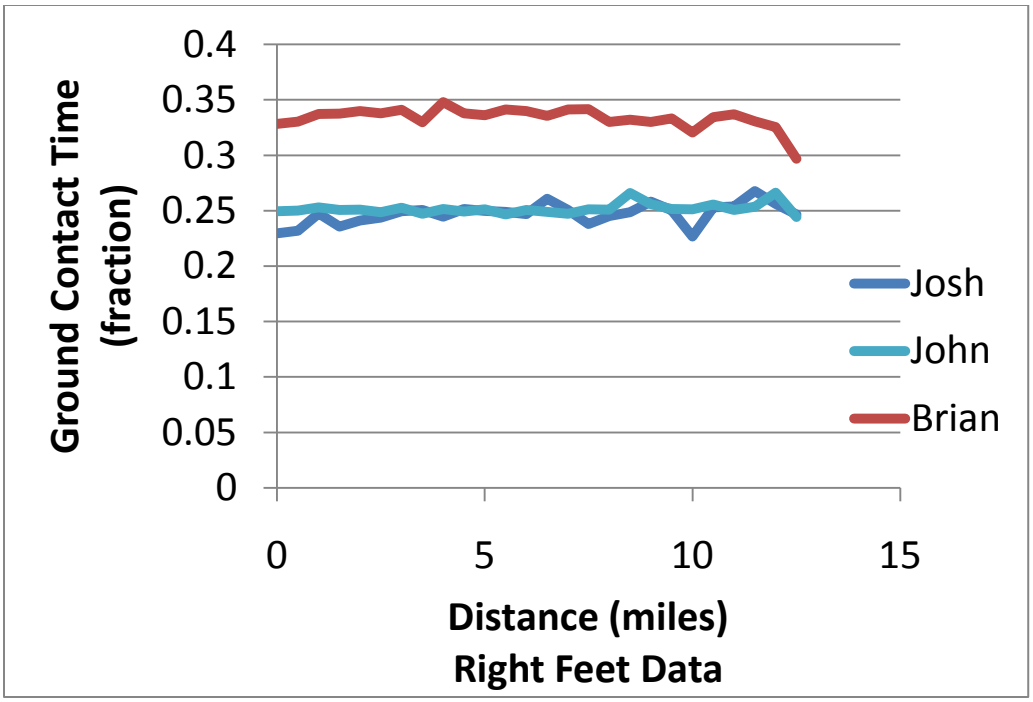


Figure 5.6: GCT_F of right feet vs. Distance for the three runners

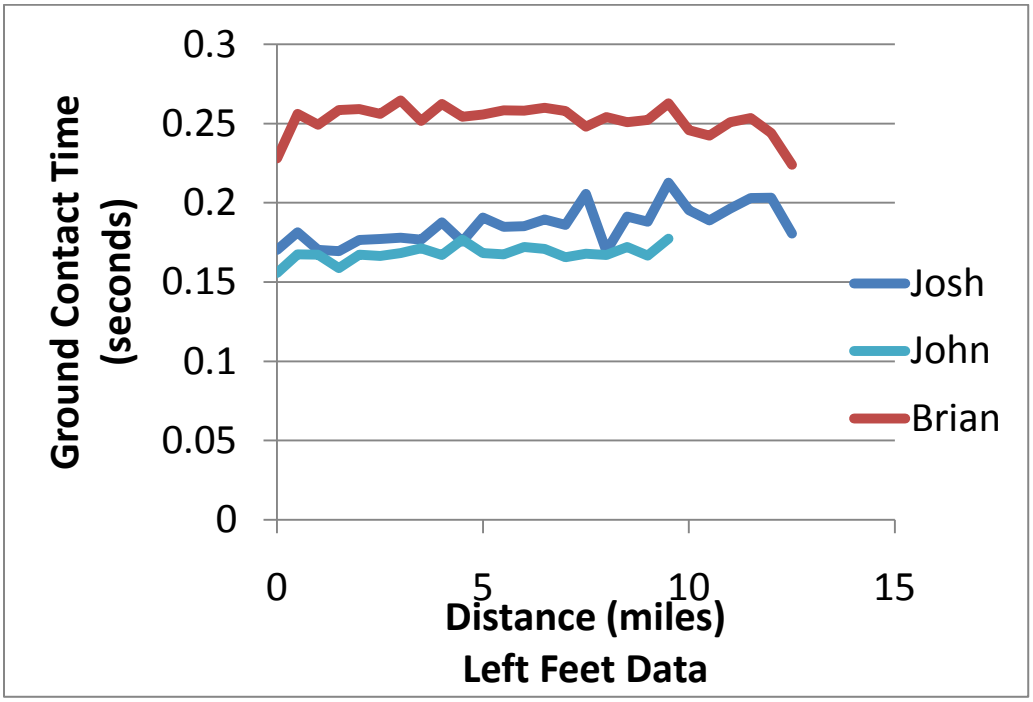


Figure 5.7: GCT_S of left feet vs. Distance for the three runners

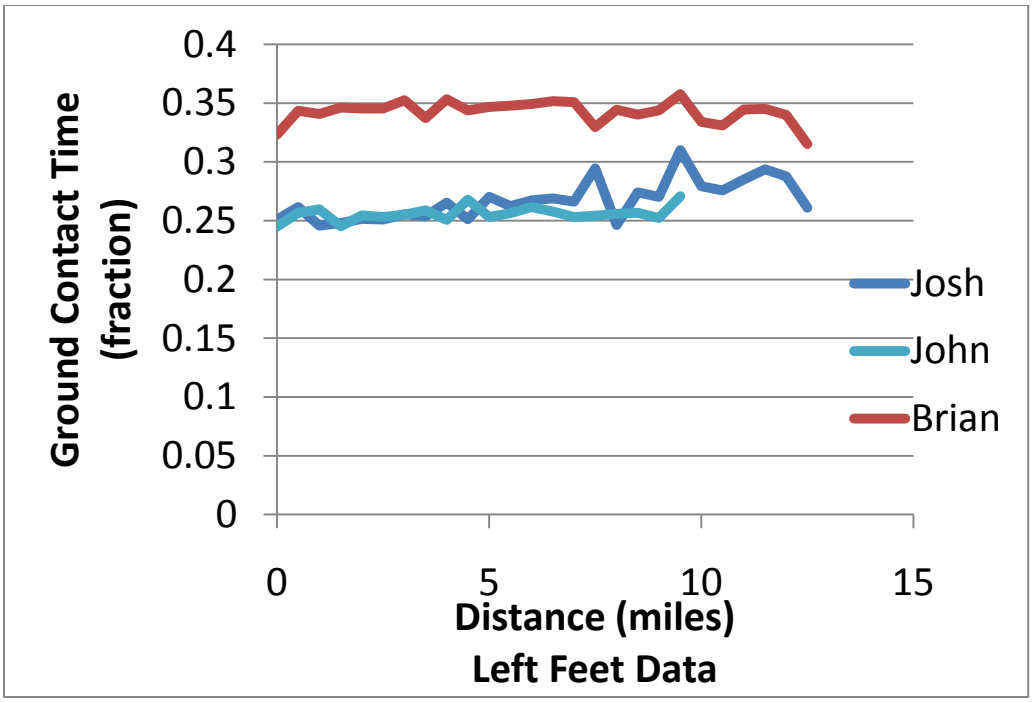


Figure 5.8: GCT_F of left feet vs. Distance for the three runners

The results are very interesting and the following can be observed from the same:

1. Both right and left foot GCT_S and GCT_F are fairly consistent for all the three runners
2. GCT_S and GCT_F values for both Josh and John are similar and considerably less than Brian's. Given that we know that elite runners have lower GCT values ([31], [28]), we could predict a much superior performance from John and Josh based on only this data. This is important from the point of the predictive capabilities of the system
3. There is a slight increase in the GCT_S values over the course of the 13.1 mile race. We attribute this to fatigue and adjustments made by the runner over the time period of the race. On the other hand, it is very interesting to note a similar effect is absent in GCT_F which remains rather unaffected over the course of the race.

This is consistent with the fact that if fatigue were the reason for increased GCT_S, it would affect every portion of the stride in equal proportions and GCT_F, thus, would remain unaffected by it

4. Finally, there is a dip in GCT_S values at the end suggesting an improved performance. This is attributed, as explained before, to the psychological effect of the finish line

5.3.4 Distal Leg Lift (DLL)

Another very important metric is the measure of the height of distal leg lift (DLL). It is a well known fact that a higher DLL leads to a better running performance. This metric has proved to highly consistent and efficient. Figure 5.9, 5.10 and 5.11 depict the time series for this metric for the three runners. Please note that the metric values are not global and hence comparing DLL values for two runners directly is not yet possible.

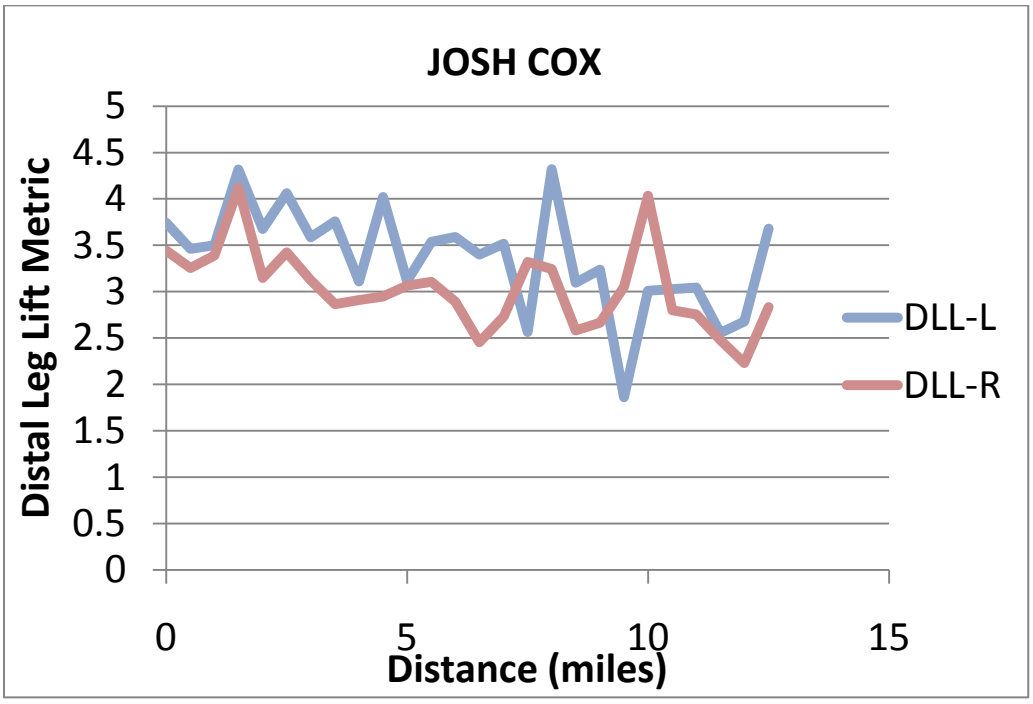


Figure 5.9: Distal Leg Lift metric with distance for Josh for both feet

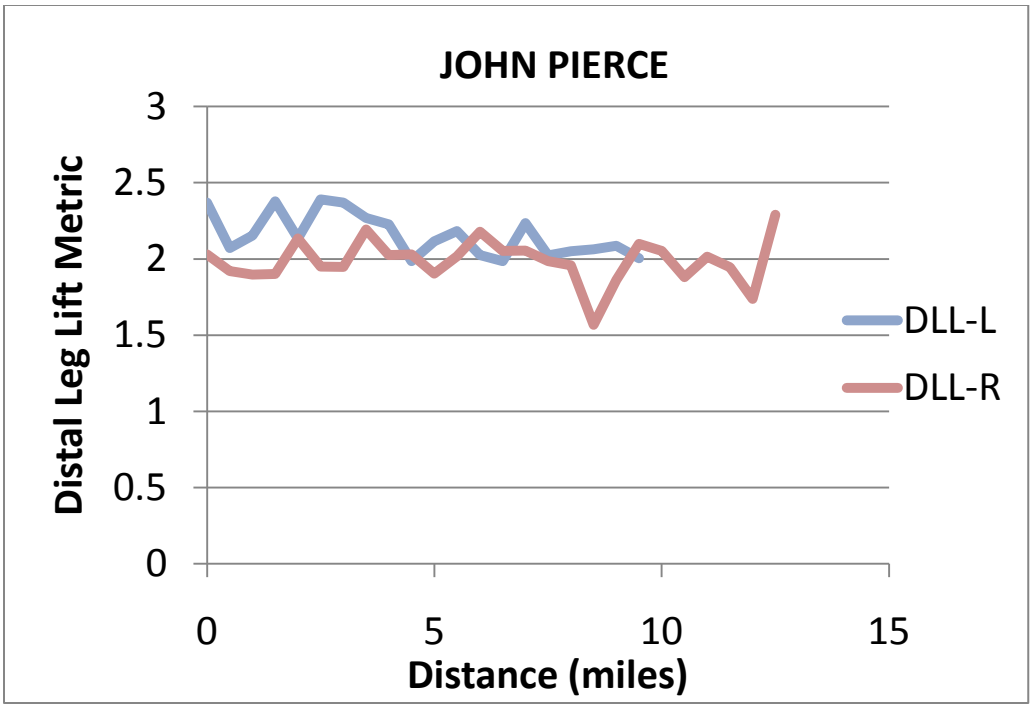


Figure 5.10: Distal Leg Lift Metric with distance for John for both feet

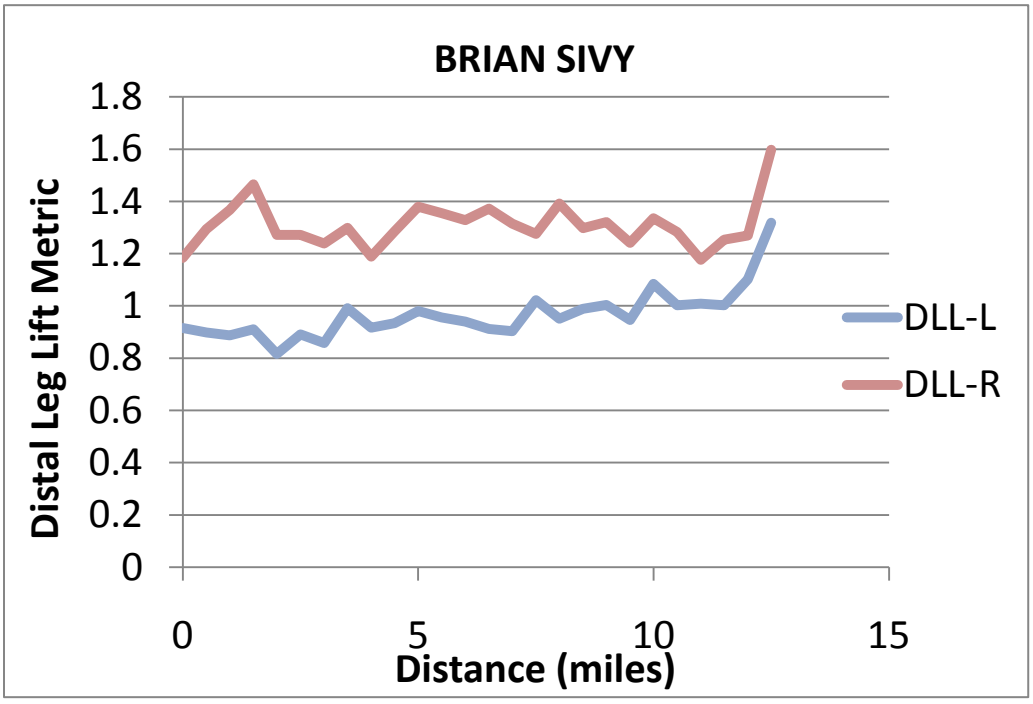


Figure 5.11: Distal Leg Lift metric with distance for Brian for both feet

Please note that DLL-L stands for the metric value for the left foot while DLL-R stands for the metric value for the right foot of the respective runners.

Following are the important observation for this metric:

1. We see that DLL values for Josh Cox and John Pierce show a decreasing trend. This is consistent with the hypothesis that fatigue might cause these changes.
2. Since higher DLL is associated with higher running performance ([4], [29]), we can conclude that performance of Josh and John dips as the race goes on. Brian's DLL values actually increase but refer to point below for explanation
3. Even though we cannot compare values close to each other since it is not a normalized metric, we can definitely draw conclusions when metric values show considerable difference. It is clear from this that Brian has a considerably lower leg lift as compared to other runners. While his performance might improve over the race period, it is not at the level of the front runners

At the time of the race, Josh was 6 feet tall and John stood at 5'8''. This explains a higher range for the DLL metric for Josh. But comparison between Josh and John is not possible at this stage of metric development. Brian stands at 6'2'' and still his metric values are considerably less than others which clearly shows his inferior DLL.

5.3.5 Over-stride Metric

Similar to the DLL metric, this metric captures the extent of the forward swing of the foot. This is designed to address the issue of over-stride. But since this metric gives a measure of the full forward swing and is not normalized yet, it is hard to succinctly state how this metric contributes to the overall running efficiency. Figure 5.12, 5.13 and 5.14 show the Over-stride metric values over the course of the race for Josh, John and Brian respectively.

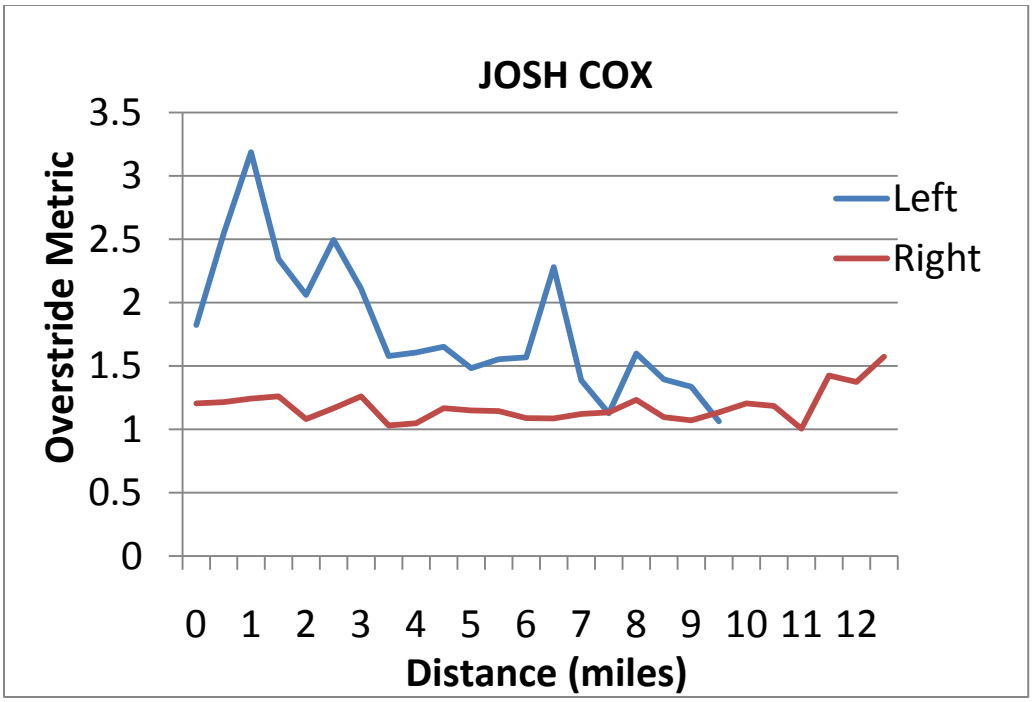


Figure 5.12: Over-stride metric vs. Distance for Josh

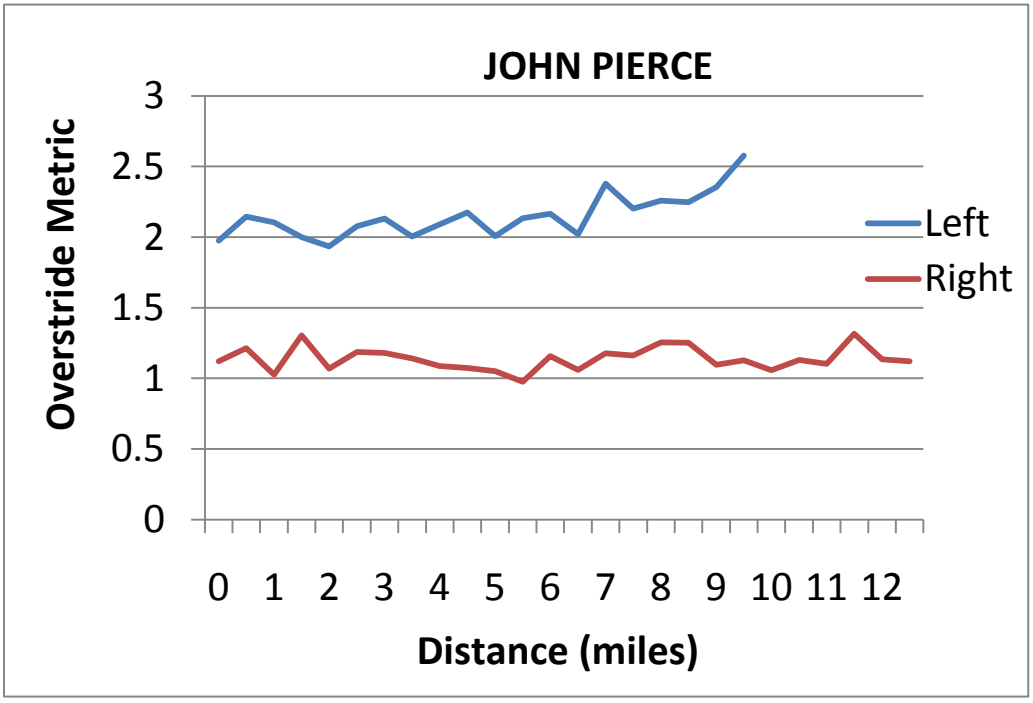


Figure 5.13: Over-stride metric vs. Distance for John

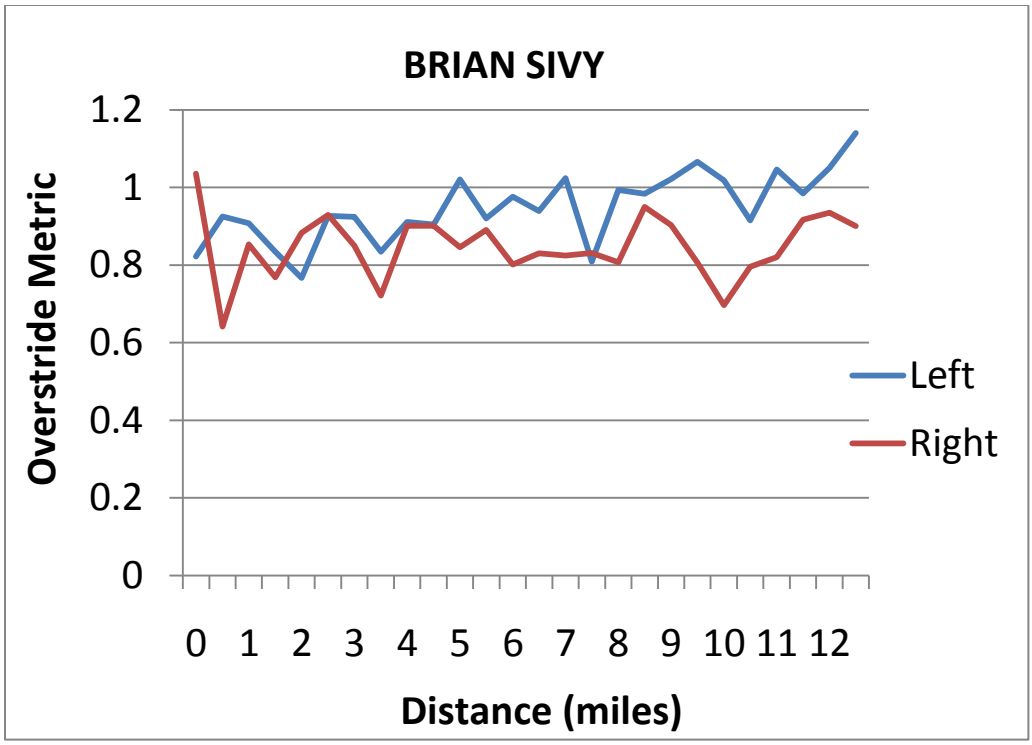


Figure 5.14: Over-stride metric vs. Distance for Brian

Following are some important observations:

1. Josh Cox has a unique running technique owing to some physical constraints. Given his unique technique, left foot forward swing is more than right foot's forward swing, which moves to the right while moving forward. This is clearly depicted in the graph. Also, fatigue is much more likely to impact his left leg swing and thus the drop in the over-stride metric
2. For John, metric shows that his left foot stride is more than that of his right foot. It was later proved using video analysis that this is indeed the case.

Hence, this metric provides key insights which are important to analyzing a runner's bio-mechanical technique.

5.3.6 Contact De-acceleration Energy Loss (CDEL)

CDEL is another important metric for analyzing running economy. Lower values of CDEL are preferred as it would mean less ground reaction forces. Please refer to section 3.3.5 for a detailed explanation for this metric. Figure 5.15, 5.16 and 5.17 give CDEL values for the race period for Josh Cox, John Pierce and Brian respectively. Please note that this metric is not yet normalized. Important observations for this metric are mentioned in the Honolulu Race results section as the effects are much more prominent there.

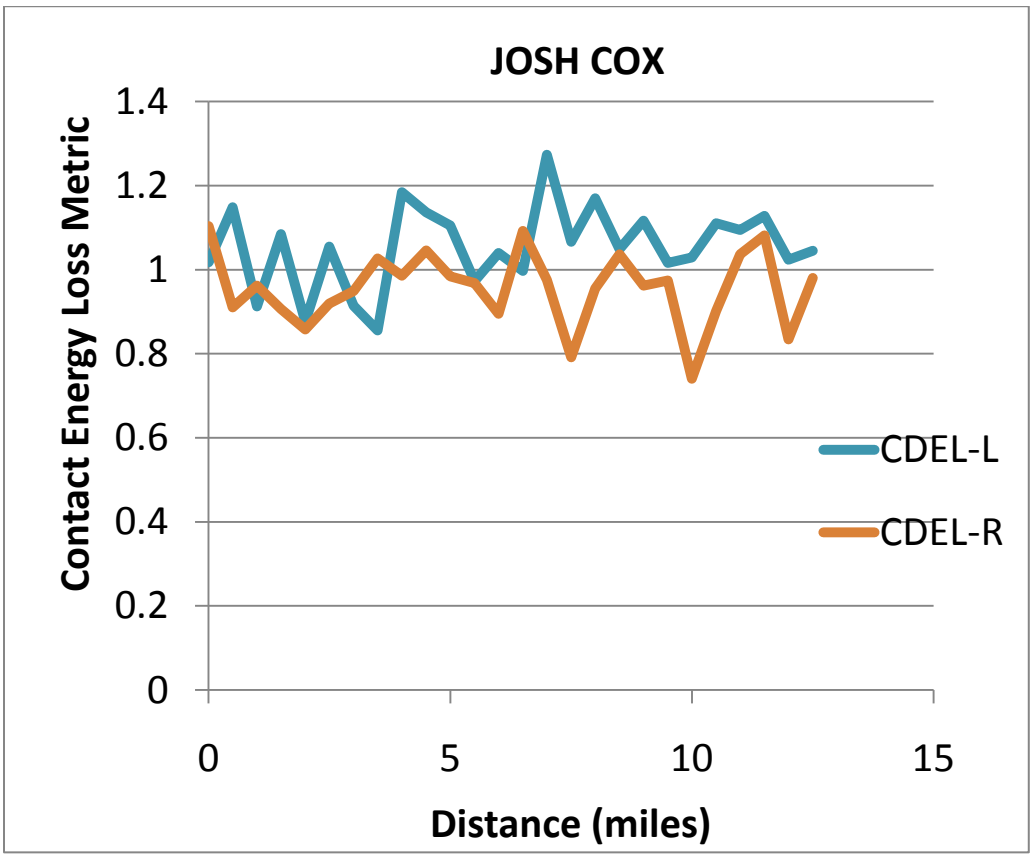


Figure 5.15: CDEL vs. Distance for Josh

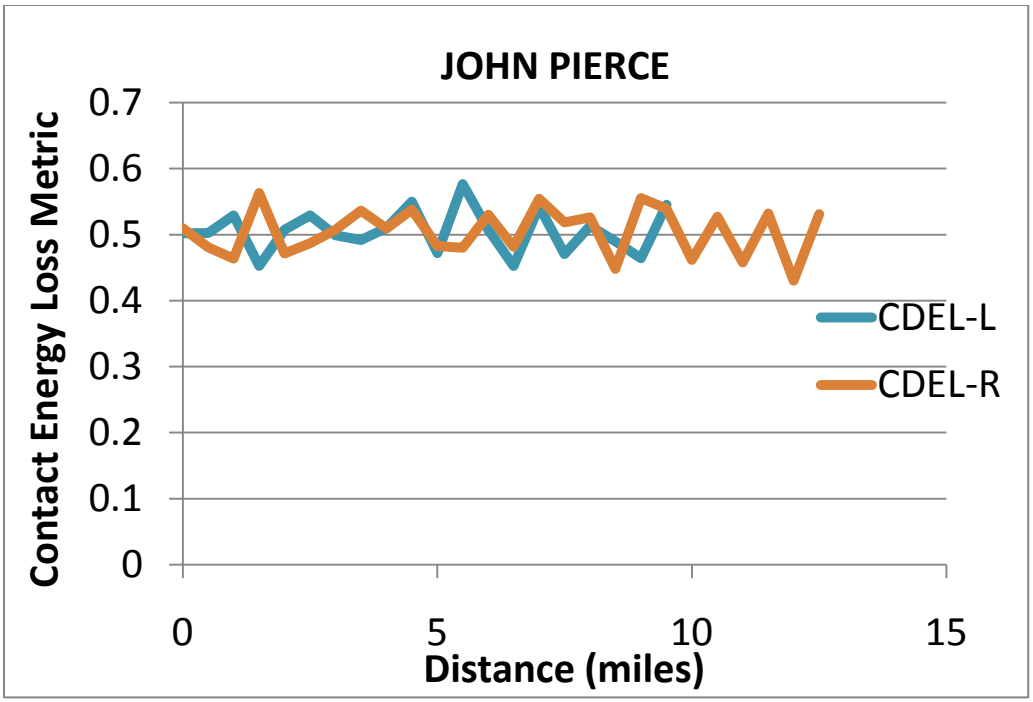


Figure 5.16: CDEL vs. Distance for John

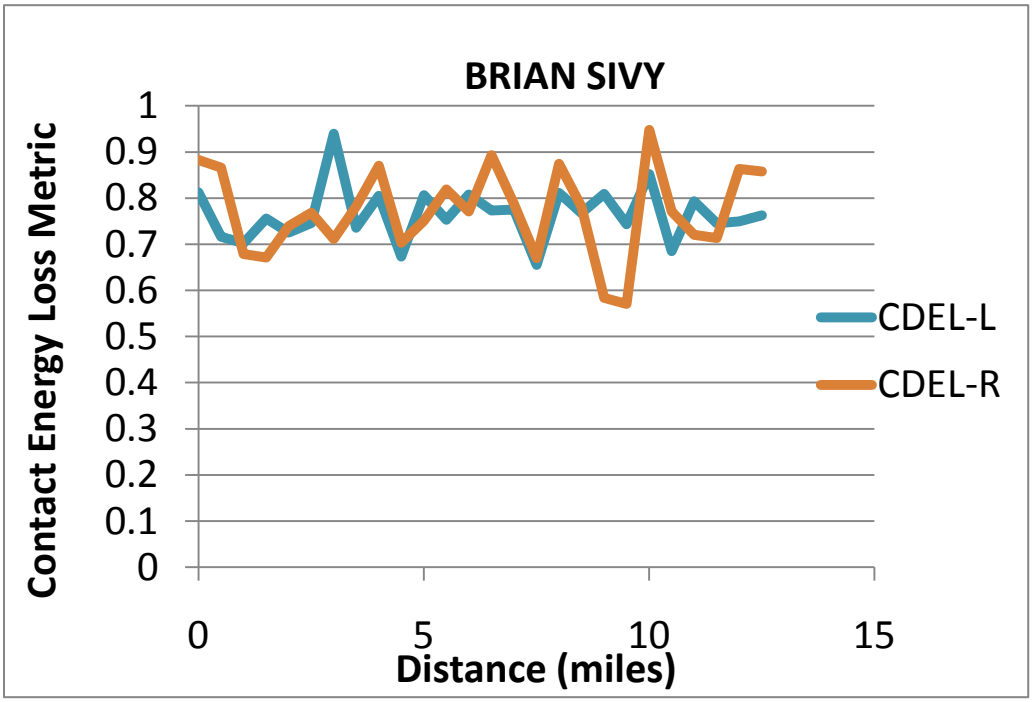


Figure 5.17: CDEL vs. Distance for Brian

5.3.7 Overall Efficiency Metric

We are able to segregate closely matched elite runners using our metrics. This is particularly exciting because this means that our system is highly sensitive to technique and performance. Differences between a non-accomplished runner and an elite runner would be alarming, a hint of which can be seen in the difference between Brian and other runners metrics. In order to sum up the performance in one single metric, we introduce here the notion of overall efficiency metric.

The most stable and reliable metrics have been cadence, GCT and DLL. We can use these to construct the overall efficiency metric. Now, a lower GCT_S implies higher efficiency and a higher DLL implies the same. Thus the relationship with these two metrics with the overall efficiency is straightforward. Exact relationship with cadence, on the other hand, is unexplored. But since all the elite runners in this race (except Brian) and in the Honolulu race presented below have a cadence in the accepted range, we will not include its formulation in the efficiency metric formulated here. Its contribution can be studied as we collect more data and the pattern becomes clearer. Hence,

$$\text{Overall Efficiency} = (\text{DLL Metric}) + (1/\text{GCT_S})$$

Please note that here the value of the DLL metric is already normalized (by a constant factor) in a way that the contribution of both metrics is equal. Because of the nature of the DLL metric, the efficiency metric is not on a scale of 1 to 10 but can go beyond. Let us now look at the efficiency of the runners during the course of the race.

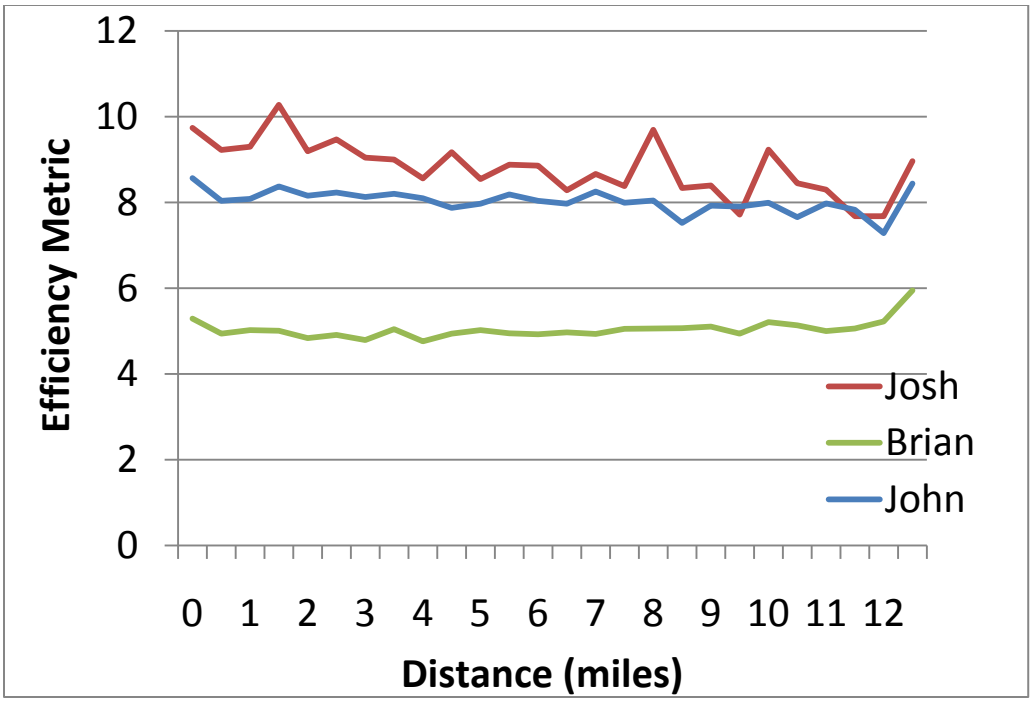


Figure 5.18: Overall Efficiency metric for the three runners

Figure 5.18 shows the efficiency metric for the three runners for the course of the race. According to this, Josh Cox performs the best followed by John closely and Brian lags behind. This has now proven to be true given the actual race results mentioned in 5.3.1.1. Also apparent is the drop in the efficiency during the course of the race which indicates fatigue. These conclusions are very important and can be potentially used to develop predictive models from the past data at any point of the race. We now move on to the next set of elite runner data we collected.

5.4 Honolulu Marathon

Following the success of the product deployment at the LA Rock and Roll Half Marathon, three more elite runners were brought on board the project and it was decided to test in the product in a full marathon this time. Following are the details of the event:

- Event: Full marathon
- Location: Honolulu, Hawaii, USA
- Distance: 26 miles (~42 Km)
- Date: 11th December 2011
- Course Elevation Chart

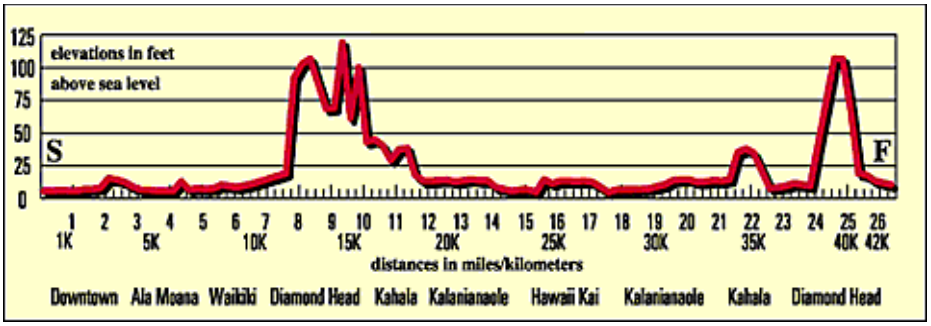


Figure 5.19: Course Elevation chart for the Honolulu Marathon

- Participating Runners: Three accomplished runners participated in the program and all of them finished in the top fifteen. Their names are as follows:
    1. Josphat Biot
    2. Jimmy Mundi
    3. Patrick Nthiwa

5.4.1 Final Race Results

Having analyzed the performance of the runners at the Honolulu Marathon, it is important to have a feedback system which confirms our findings and analysis. Actual race performance is a very good indicator for the same. Figure 5.20 shows the final race results for the three runners. It is worth mentioning here that while some of the metrics were refined using LA Rock and Roll Half Marathon, no changes were required in the metric designs for the Honolulu Marathon.

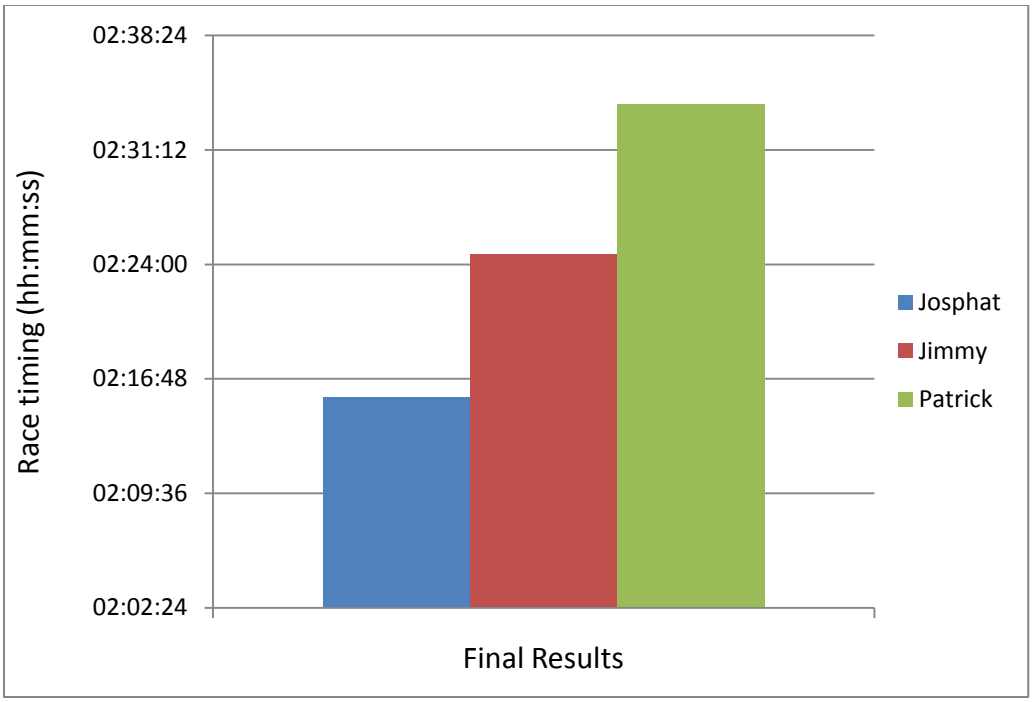


Figure 5.20: Final race timings and comparison of the three runners

As can be seen, Josphat Biot performed the best and finished 3rd with a timing of 02:15:40. He was followed closely by Jimmy Mundi at 6th position with Patrick Nthiwa finishing 15th. All the three runners were close and it would be interesting to observe if the metrics can distinguish between them. Please note that owing to some technical glitches, minor portions of Patrick's data were corrupted and this can be seen by negative spikes in the graphs. These bear no consequence to the actual metrics.

Since this was a full marathon, we expect certain effects to be pronounced for certain metrics. Hence it is also important to see the actual race performance with time and not just at the finish line. Figure 5.21 shows comparisons at 10 Km, at the half way stage (21 Km) and at the finish line. Clearly all the three runners were neck and neck at the half way stage, after which the performance of Jimmy and Patrick dropped while Josphat kept up his pace.

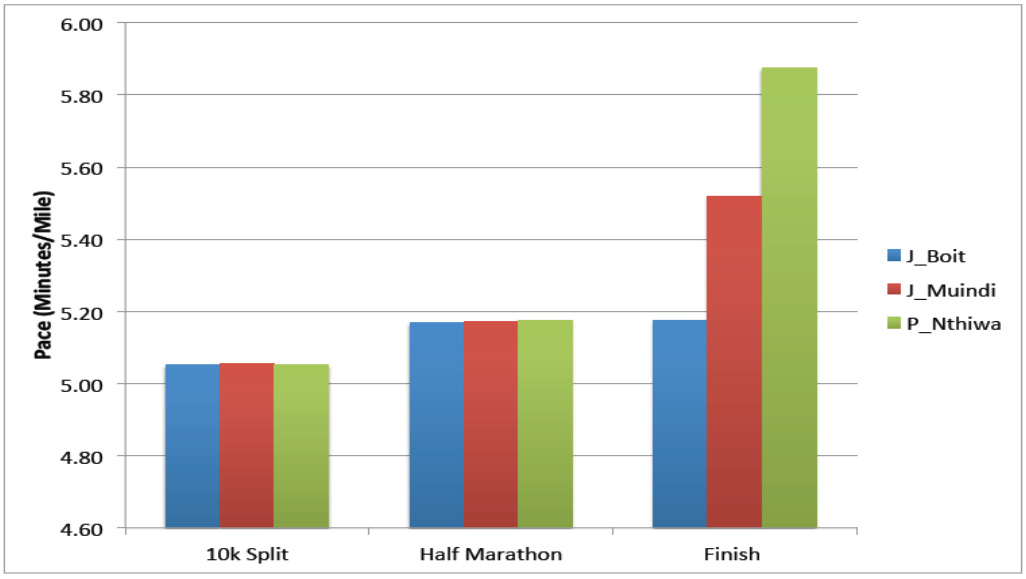


Figure 5.21: Split timing comparison for the three runners

5.4.2 Cadence

Let us first examine the cadence of all the three runners over the length of the entire race.

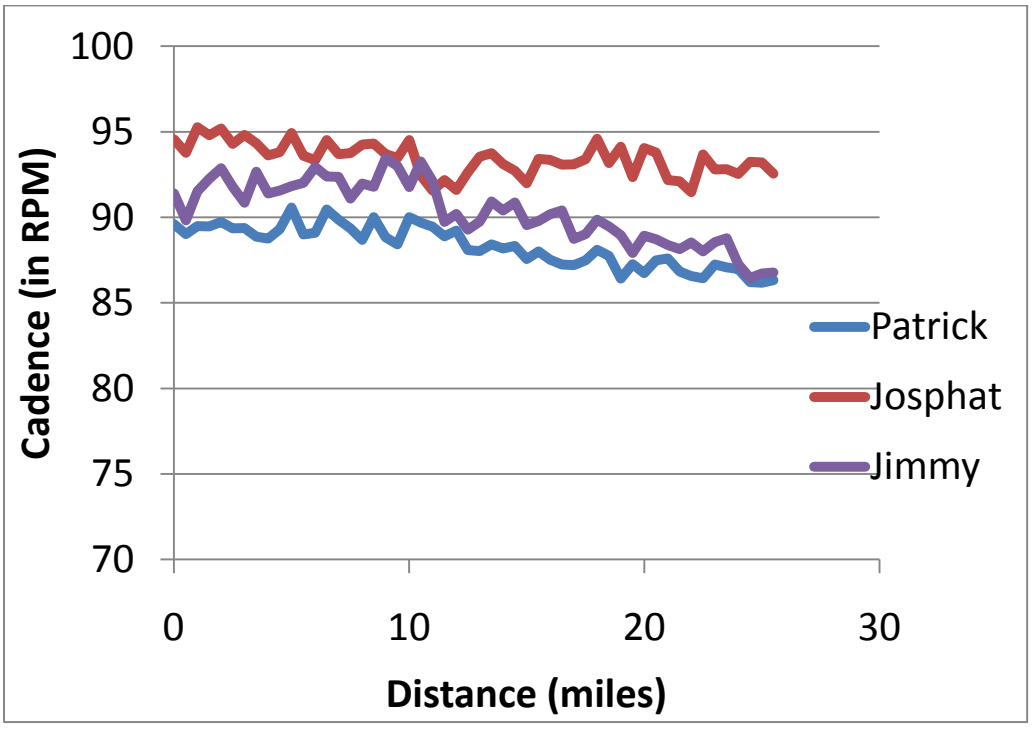


Figure 5.22: Distance vs. Cadence time series for the three runners

Figure 5.22 gives the cadence values for the entire race for all the three runners. A few preliminary observations and conclusions from the graphs above are:

1. Cadence lies in the range of 85-95 rpm at all times during the race for the three runners. While Josphat's cadence is on high end of the spectrum, Jimmy and Patrick begin close to 90
2. Josphat's cadence drops only slightly over the course of the race while Jimmy and Patrick experience considerable drop in their values after the half way stage. This is consistent with the actual pace of the runners plotted in Figure 5.21

Again the ability of this metric to closely track race performance is an asset and thus is a very good indicator of the running efficiency.

5.4.3 Ground Contact Time (GCT)

Refer to Figure 5.23, 5.24, 5.25 and 5.26 to see the time series graph for all the three runners for GCT_S and GCT_F metrics (left and right foot). Following can be observed and/or concluded from the results:

1. Values of these metrics are fairly close for the three runners. Upon careful examination, it is apparent that Josphat's GCT values are the lowest
2. As the race progresses, GCT values increase in magnitude. This signals a drop in performance. The fact that this happens gradually over the course of the race clearly suggests fatigue. While fatigue plays a role for all the runners, it less prevalent in Josphat's data as compared to Jimmy and Patrick

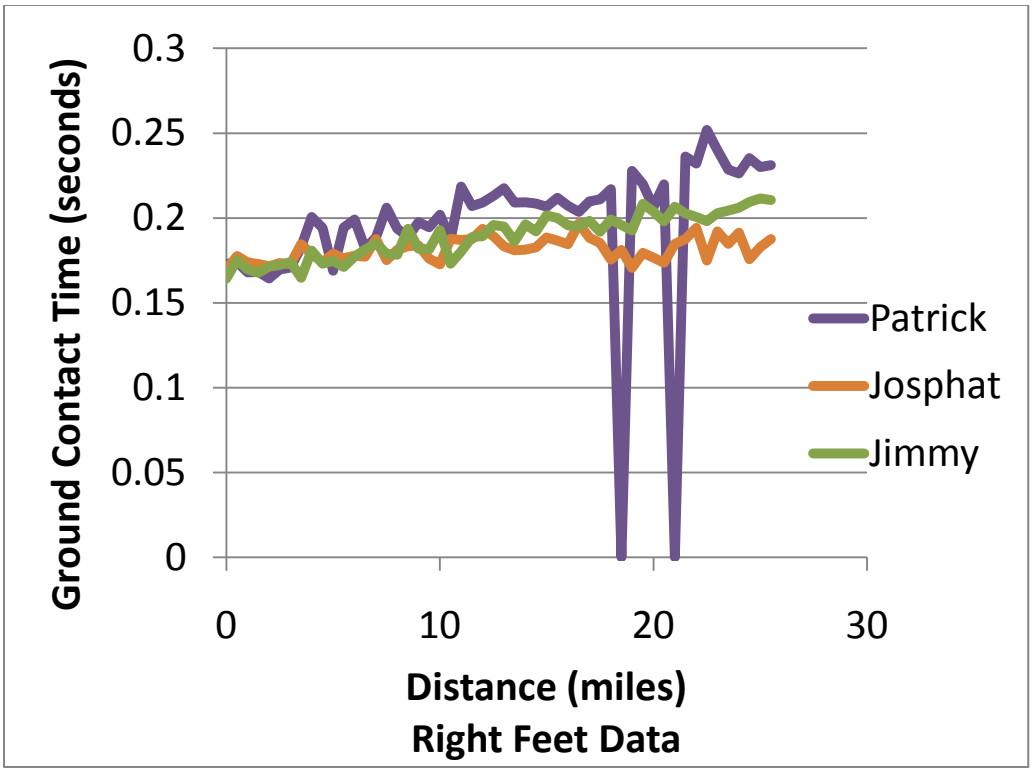


Figure 5.23: GCT_S of right feet vs. Distance for the three runners

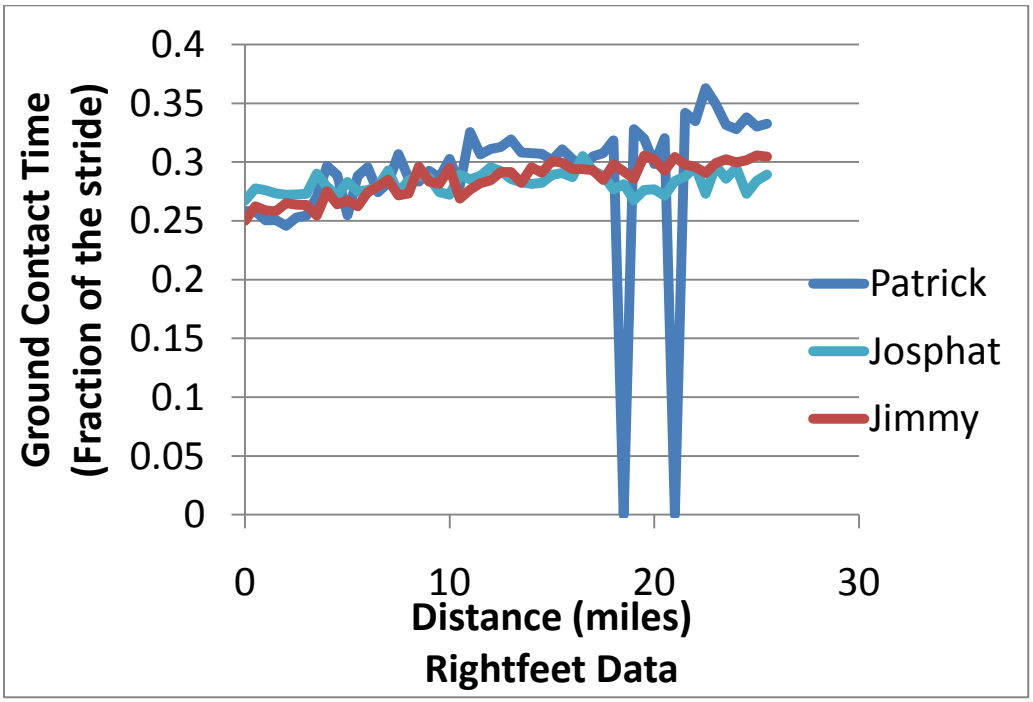


Figure 5.24: GCT_F of right feet vs. Distance for the three runners

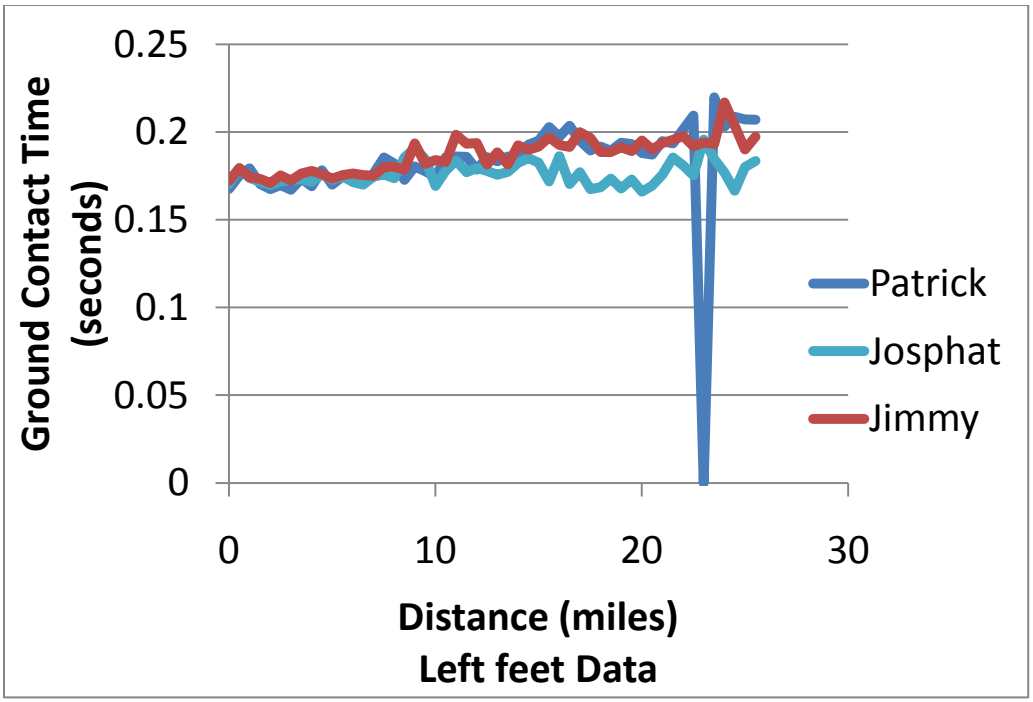


Figure 5.25: GCT_S of left feet vs. Distance for the three runners

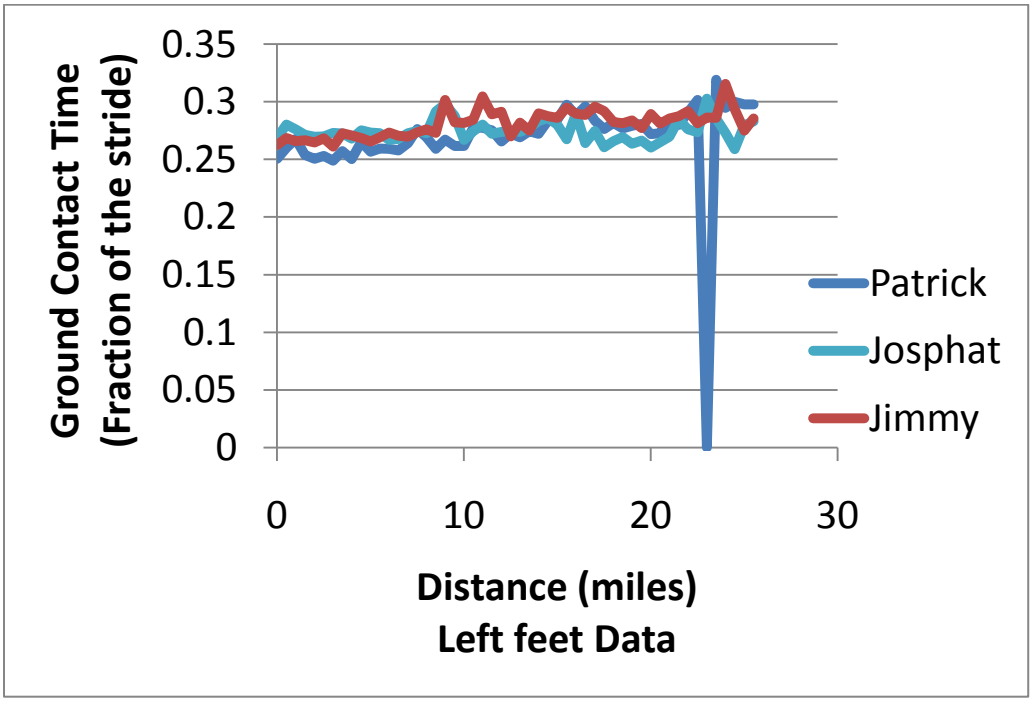


Figure 5.26: GCT_F of left feet vs. Distance for the three runners

Both of the above observations align with the actual results. While initially all the runners were close, Jimmy and Patrick experienced a drop in the performance as the race went on. This is highlighted in Figure 5.21. Hence this metric is highly valuable for performance monitoring in actual race conditions.

5.4.4 Distal Leg Lift (DLL)

Please recall that a higher DLL leads to a better running performance. Figure 5.27, 5.28 and 5.29 depict the time series for this metric for Josphat, Jimmy and Patrick respectively. Please note that though the metric values are not global and are dependent on the height of the runner, in this case the height of all the three runners is comparable (within one inch of each other) and hence we can compare the runner performance for this metric. Following are important observation for this metric:

1. Josphat's DLL metric values remain rather constant throughout the race; values for Jimmy and Patrick experience considerable dip over the race period. Hence it can be safely concluded that fatigue plays a big role in the performance dip for Jimmy and Patrick. This is in accordance with the race results highlighted in Figure 5.21.
2. Josphat's metric value is around 2.25 on average while it's around 1.5 for Jimmy. This indicates that the extent of distal leg lift is more for Josphat than Jimmy. This indeed turned out to be the case upon careful video analysis. Thus, this metric highly reliable and efficient.

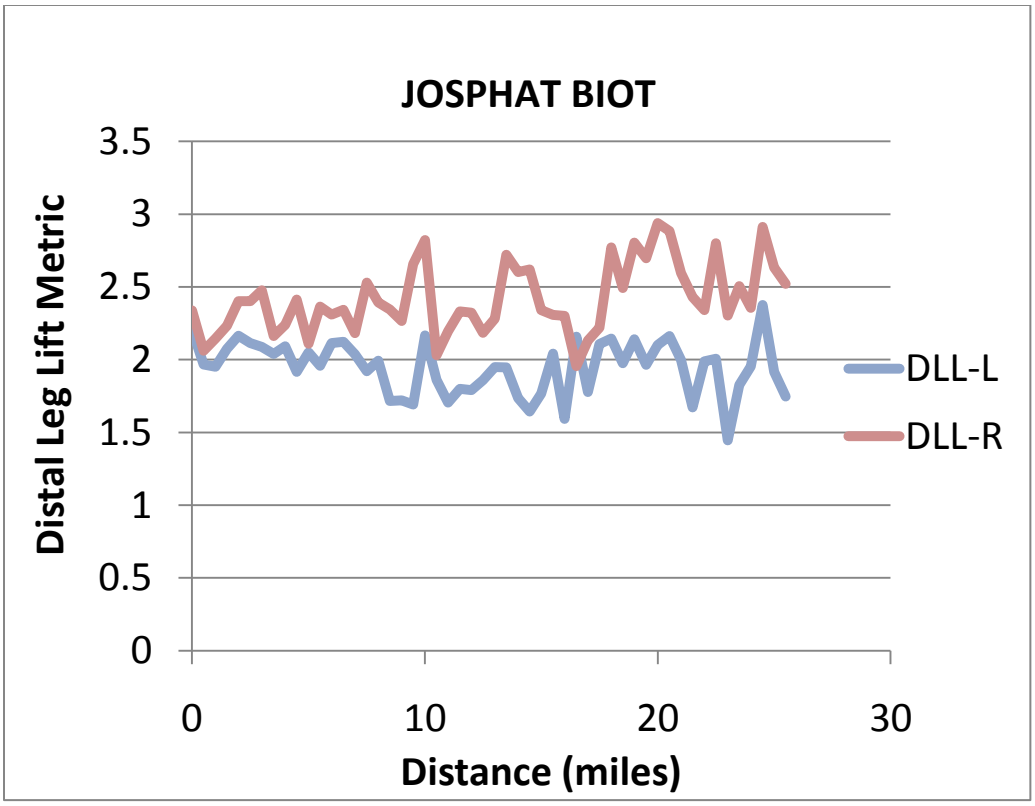


Figure 5.27: Distal Leg Lift metric with distance for Josphat for both feet

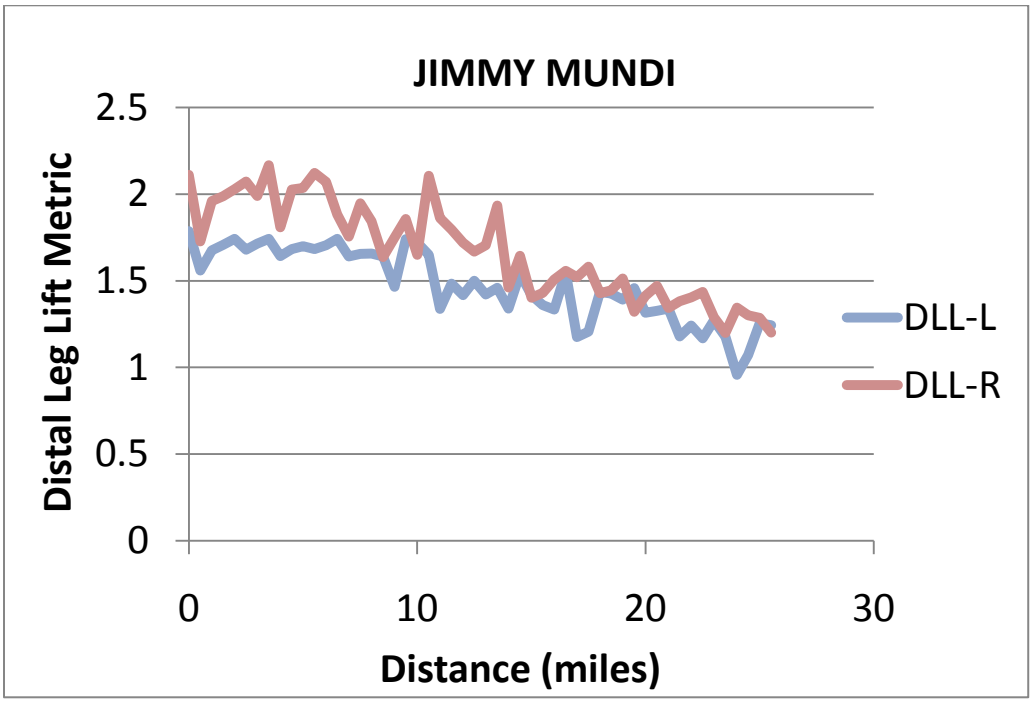


Figure 5.28: Distal Leg Lift metric with distance for Jimmy for both feet

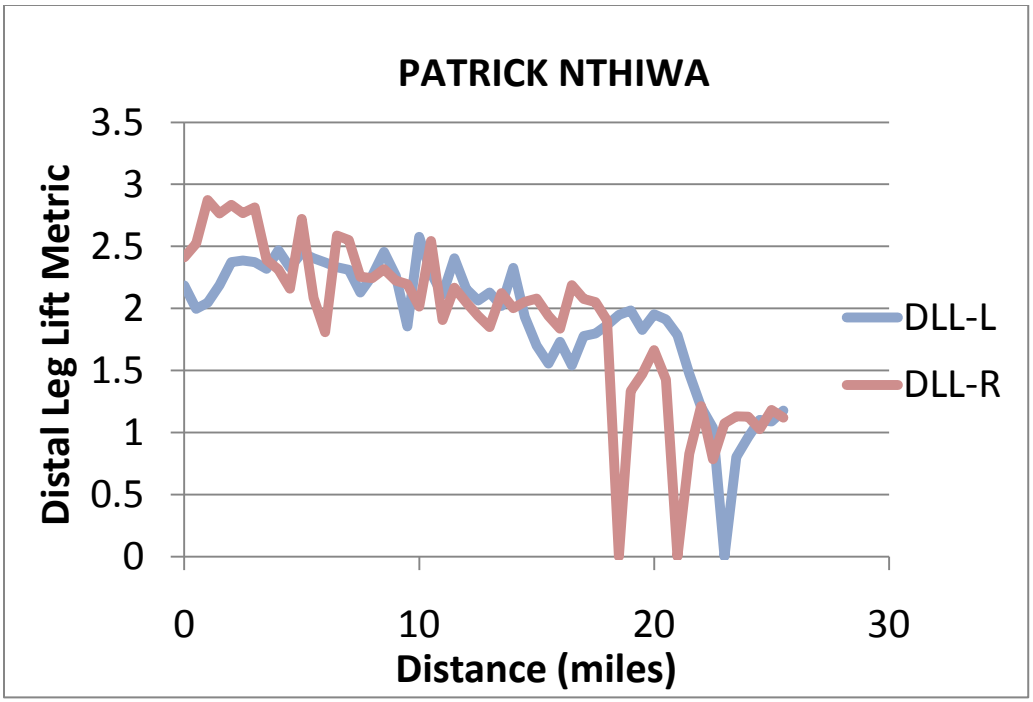


Figure 5.29: Distal Leg Lift metric with distance for Patrick for both feet

Please note that DLL-L stand for DLL metric value of the left foot and DLL-R stands for the DLL metric value of the right foot for the respective runners.

5.4.5 Over-stride Metric

Similar to the DLL metric, this metric captures the extent of the forward swing of the foot. This is designed to address the issue of over-stride. Figure 5.30, 5.31 and 5.32 show the Over-stride metric values for Josphat, Jimmy and Patrick respectively. From the graphs it is apparent that the extent of forward swing remains consistent for Josphat while it reduces with time for Jimmy and Patrick. Also, the differences in the left and right foot metric values for Jimmy are interesting to note. Jimmy's left foot stride probably suffers from a case of minor under-stride. This metric thus provides key insights which are important to analyzing a runner's bio-mechanical technique.

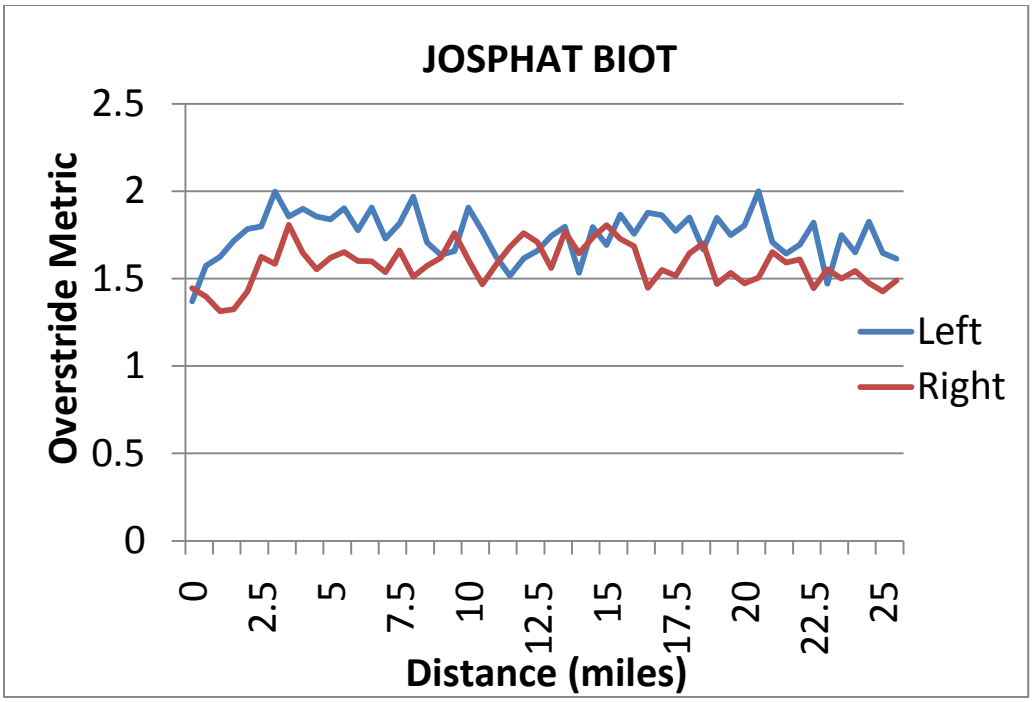


Figure 5.30: Over-stride metric vs. Distance for Josphat

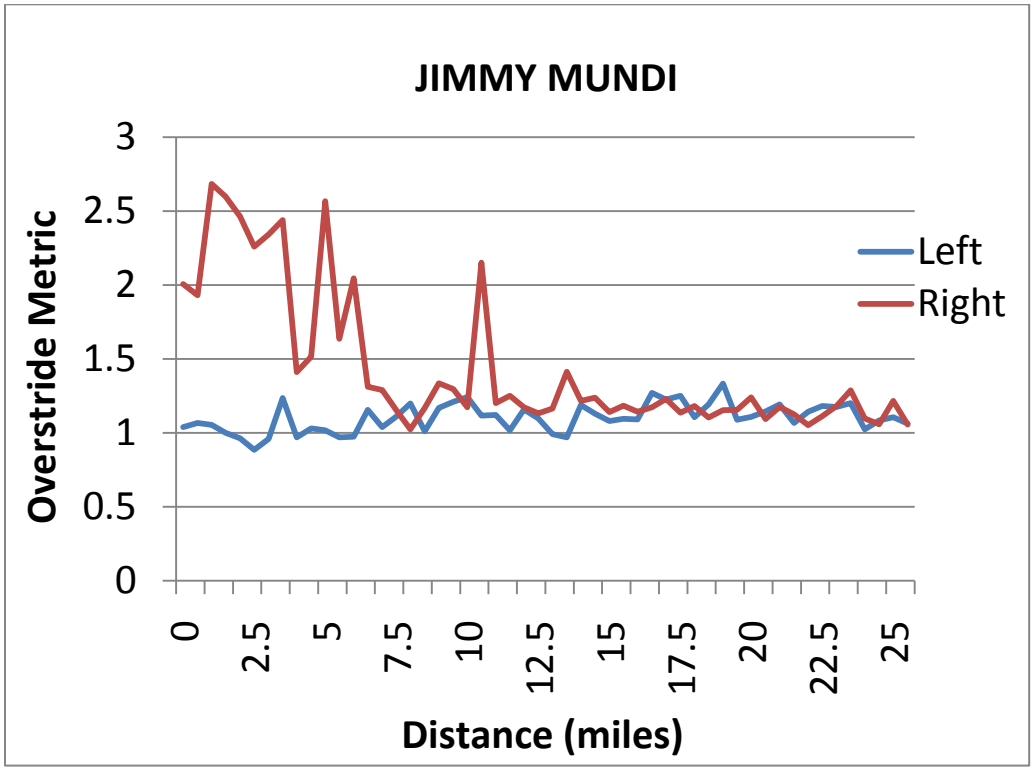


Figure 5.31: Over-stride metric vs. Distance for Jimmy

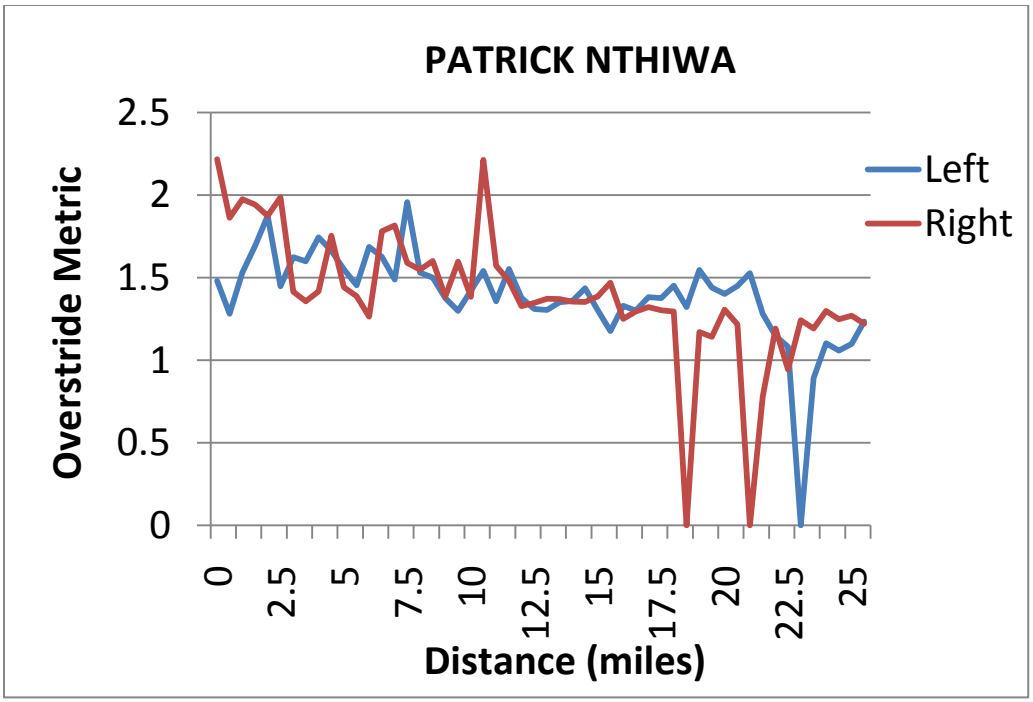


Figure 5.32: Over-stride metric vs. Distance for Patrick

5.4.6 Contact De-acceleration Energy Loss (CDEL)

CDEL is another important metric for analyzing running economy. Figure 5.33, 5.34 and 5.35 give CDEL values for the race period for Josphat, Jimmy and Patrick respectively. From the graphs below, no specific trends emerge as a function of time. While the values increase with time for Josphat, they decrease for Jimmy and remain rather steady for Patrick. This metric is a function of the weight of the runner. Given the trends, it can be concluded that different runners respond differently to fatigue as far as breaking energy loss is concerned. While some might land with a bigger 'thump' owing to fatigue (probably like Josphat), others reduce the stride length with time leading to a natural decrease in the breaking energy loss. The insights provided are nonetheless important and are an important part of the bio-mechanical technique analysis.

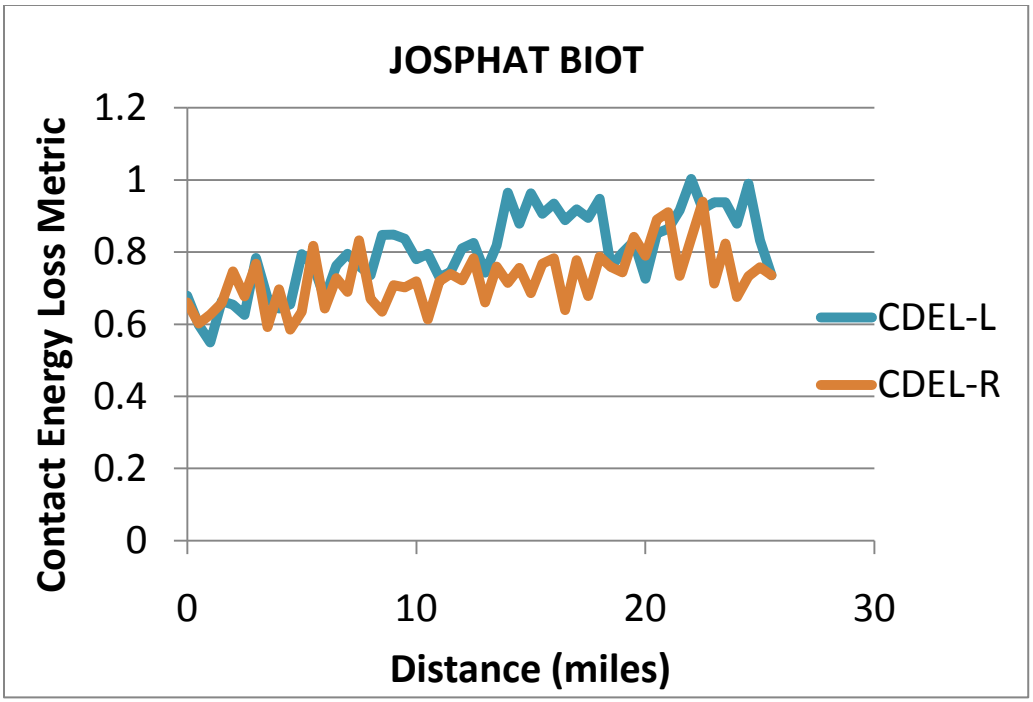


Figure 5.33: CDEL metric vs. Distance for Josphat

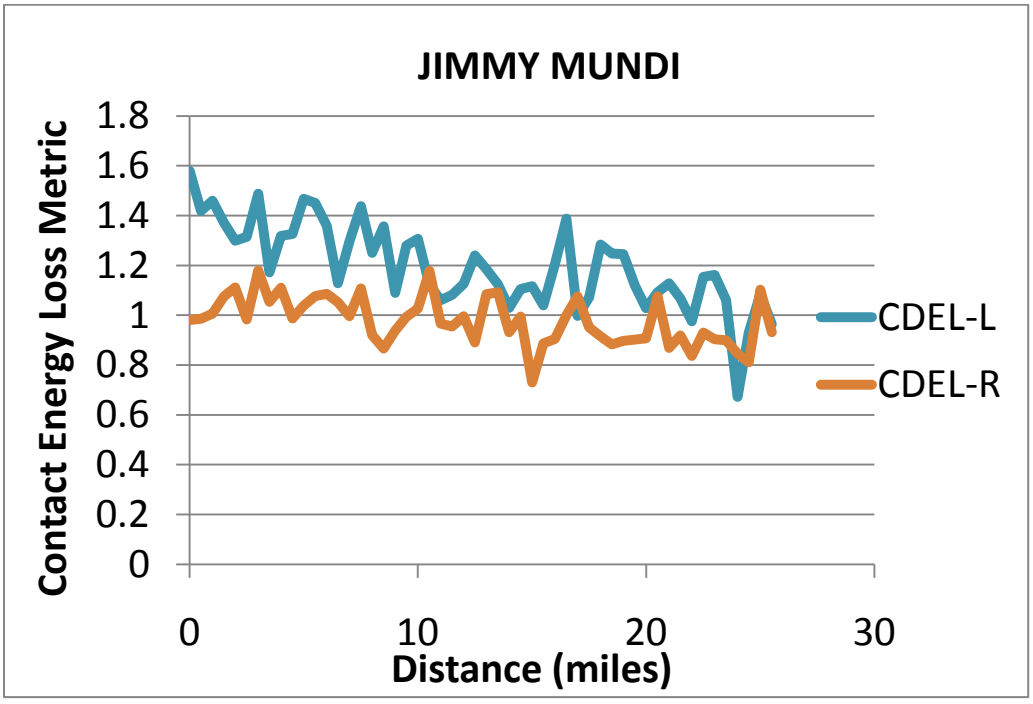


Figure 5.34: CDEL metric vs. Distance for Jimmy

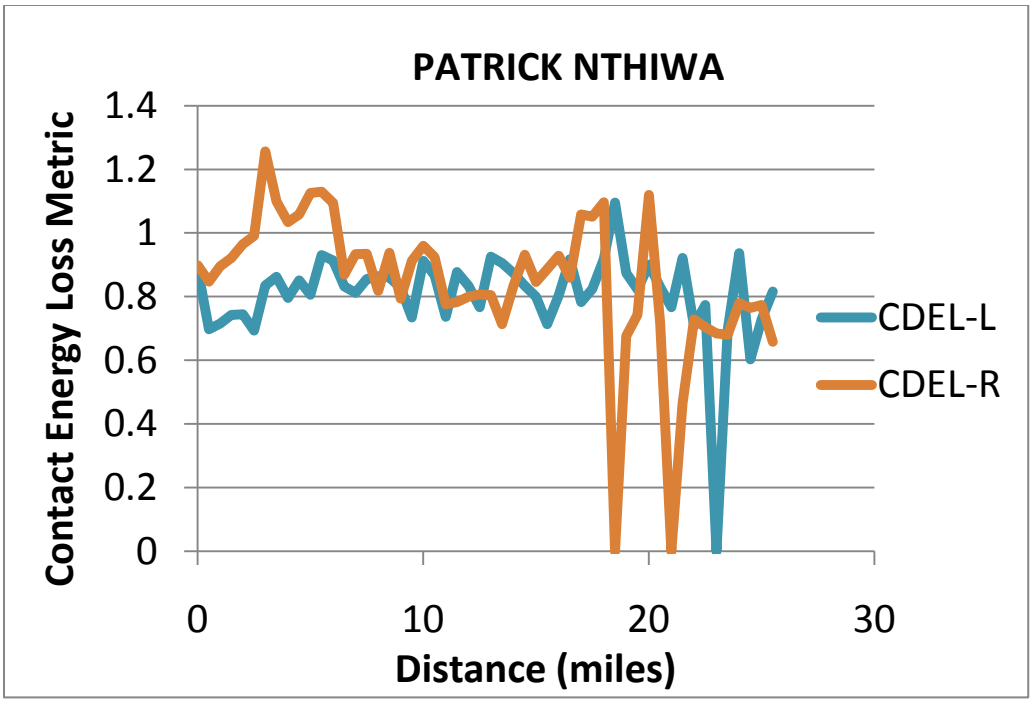


Figure 5.35: CDEL metric vs. Distance for Patrick

5.4.7 Overall Efficiency Metric

As we saw in the analysis of the LA Rock and Roll Marathon, the efficiency metric was able to track the performance of each runner and reflected the true performance of the runners. We now check if this holds true for the Honolulu marathon results as well. Figure 5.36 shows the same metric results for Josphat, Jimmy and Patrick.

As can be seen from Figure 5.36, Josphat performs better than Jimmy who in turn performs slightly better than Patrick. We know from actual results that this is true and that all the three runners are closely matched.

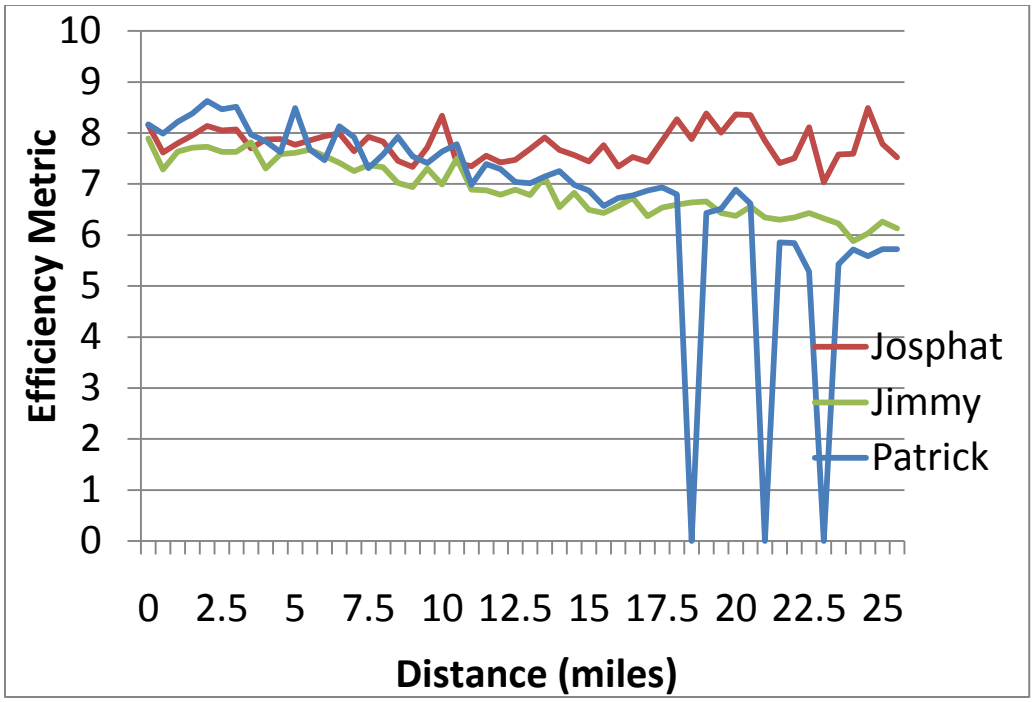


Figure 5.36: Overall Efficiency metric with Distance for the three runners

It is even more interesting to observe are the race splits. The race splits depicted in section 5.3.2.1 show that initially all the runners were almost at par. This is clearly visible here and it only after the 10 mile point that we are able to segregate the performance of these three elite athletes. While Josphat maintains his performance, that of Jimmy and Patrick drops from thereon in.

5.5 Other Important Results

While many important results and patterns were verified and examined using the race data, we still need to examine the metric value dependence on pace. While fatigue and running technique played a huge role in determining the outcome of the marathons, markers are needed to differentiate between long distance running and shorter runs. For this purpose and for collecting more data and creating a database, trial runs were conducted with three professional athletes at the Olympic Training Center (OTC) at

Mammoth, CA, USA. The trial runs included 200m, 300m, 400m runs and a tempo run by one of the athletes. Following athletes participated in the trial runs:

1. Angela Bizzari
2. Morgan Uceny
3. Anna Pierce

Let us now examine some of the results in brief from the trial data for the three runners.

5.5.1 Angela Bizzari

Angela ran six 1 Km runs with rest after each run. The resting time was approximately equal to the run time but it is just represented by just a break in the data series in the figures below. The data presented here examines the GCT_S and the DLL metric for the purposes of brevity. These are represented by Figure 5.37 (GCT_S) and Figure 5.38 (DLL). Please note that the break in the graph represents a break between consecutive runs.

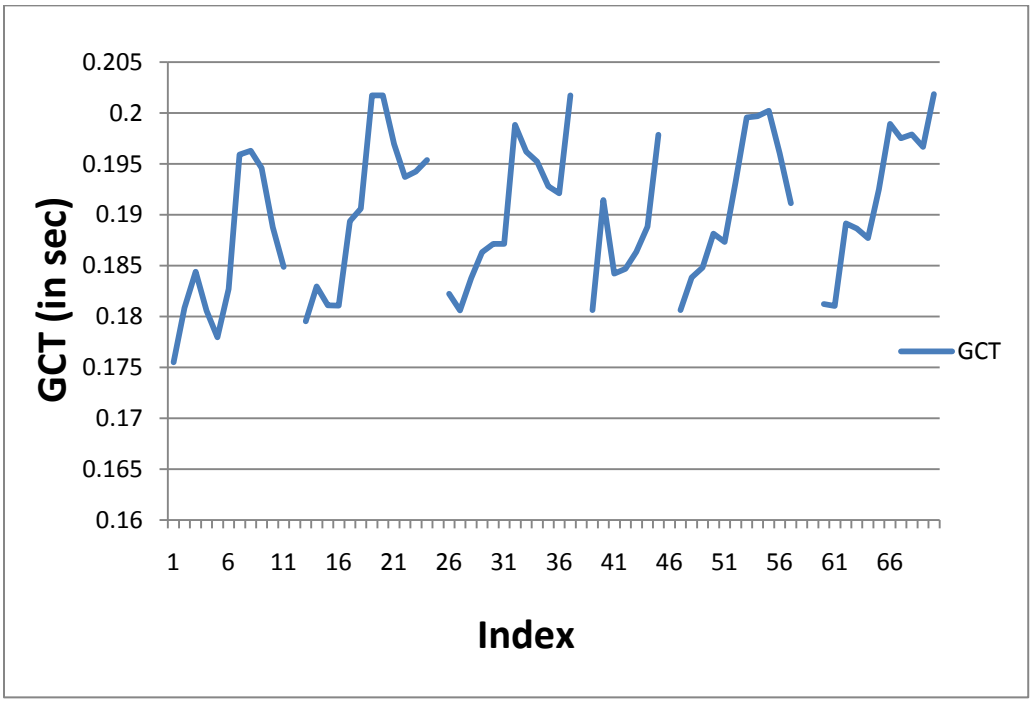


Figure 5.37: Progression of GCT_S for the six 1K runs by Angela

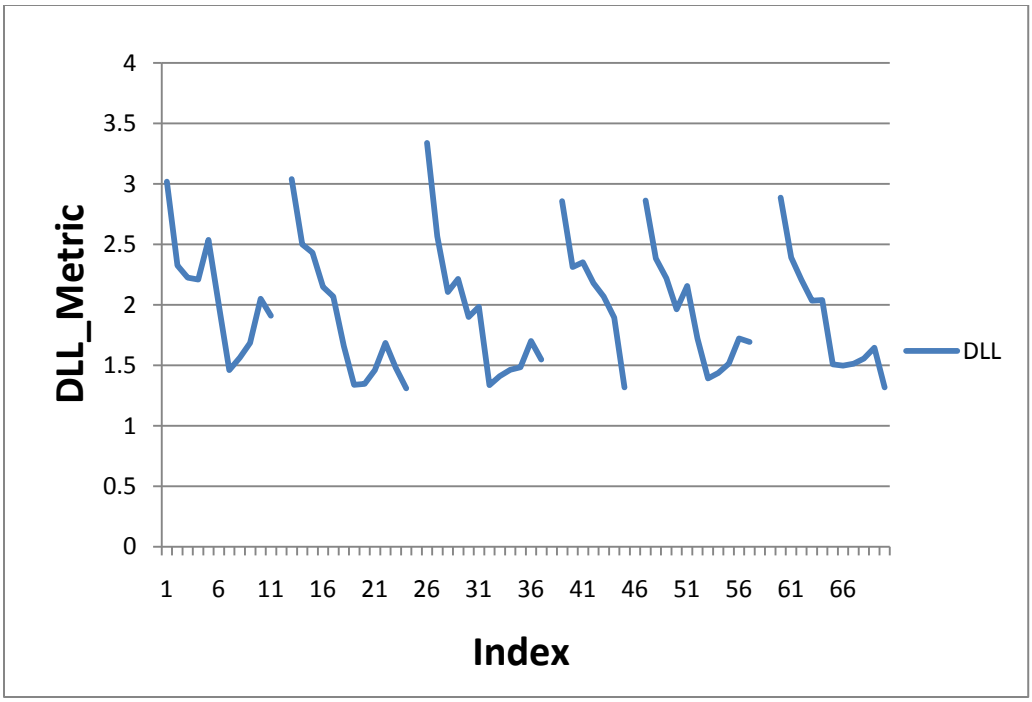


Figure 5.38: Progression of DLL metric for the six 1K runs by Angela

It is very interesting to note that during the course of each run, there is a remarkable drop in the DLL and the GCT_S metric values. This clearly indicates fatigue and a change of technique as the run goes along. Also, after the resting period and at the beginning of each run, we see a reset of values which suggests that recuperating effects are in play. Also, as we move from the first run to the sixth one, we can see that the starting metric values for each run are not the same. This suggests a need for a better training regime for the runner so that the change in the running bio mechanics is not so drastic over the course of an actual race. As noted before, increase in GCT and decrease in the DLL metric values lead to reduced running efficiency.

5.5.2 Morgan Uceny

Morgan trained for three 400m runs, four 300m runs and a tempo run at the Olympic Training Center. Figure 5.39 and 5.40 give the GCT_S and GCT_F metric graphs.

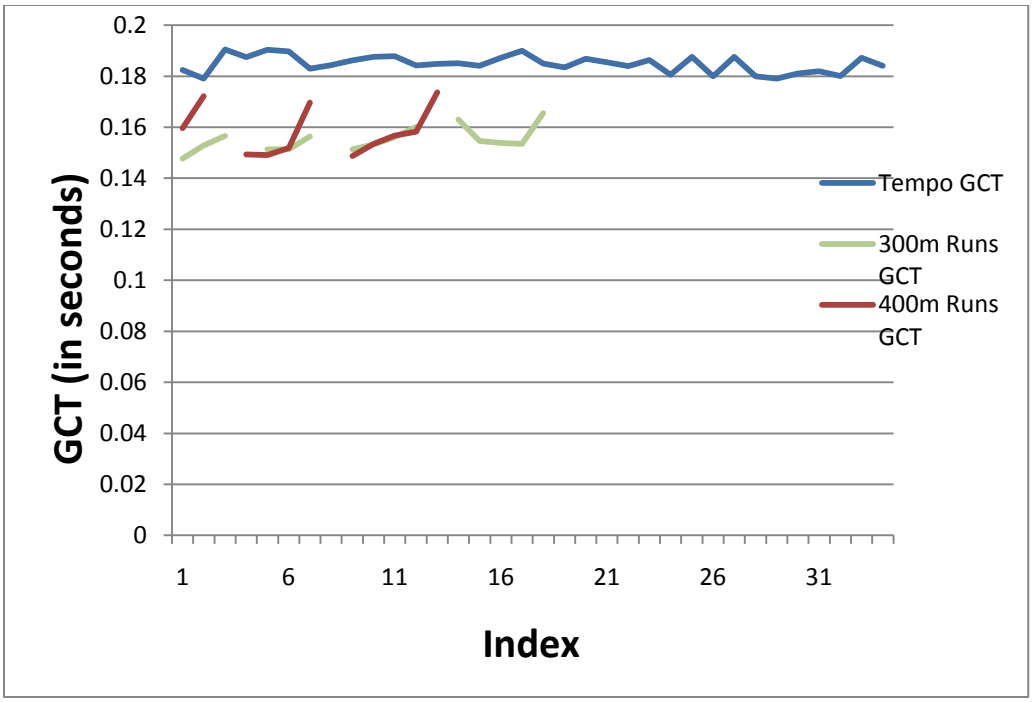


Figure 5.39: Progression of GCT_S for the different runs by Morgan

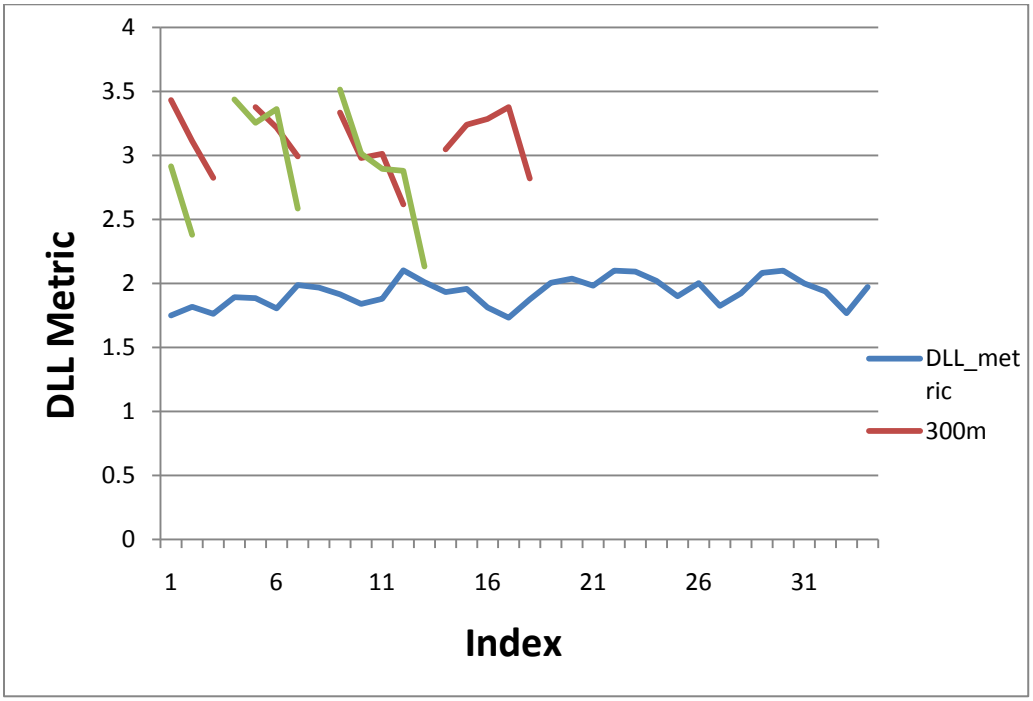


Figure 5.40: Progression of DLL metric for the different runs by Morgan

While individual run results are as expected and follow the trend established by other runner metrics, it is interesting to observe the metric values for the tempo run. While the GCT is considerably higher, the DLL_metric is lower. This points out that the runner follows a relaxed approach while doing the tempo run with lower leg lift values. Hence efficiency is set to drop. It is important here to note that some of these effects might be a factor of pace as well.

5.5.3 Anna Pierce

Anna completed three 400m runs, eight 200m runs and five 300m runs. Important metric values are highlighted in Figure 5.41 and Figure 5.42.

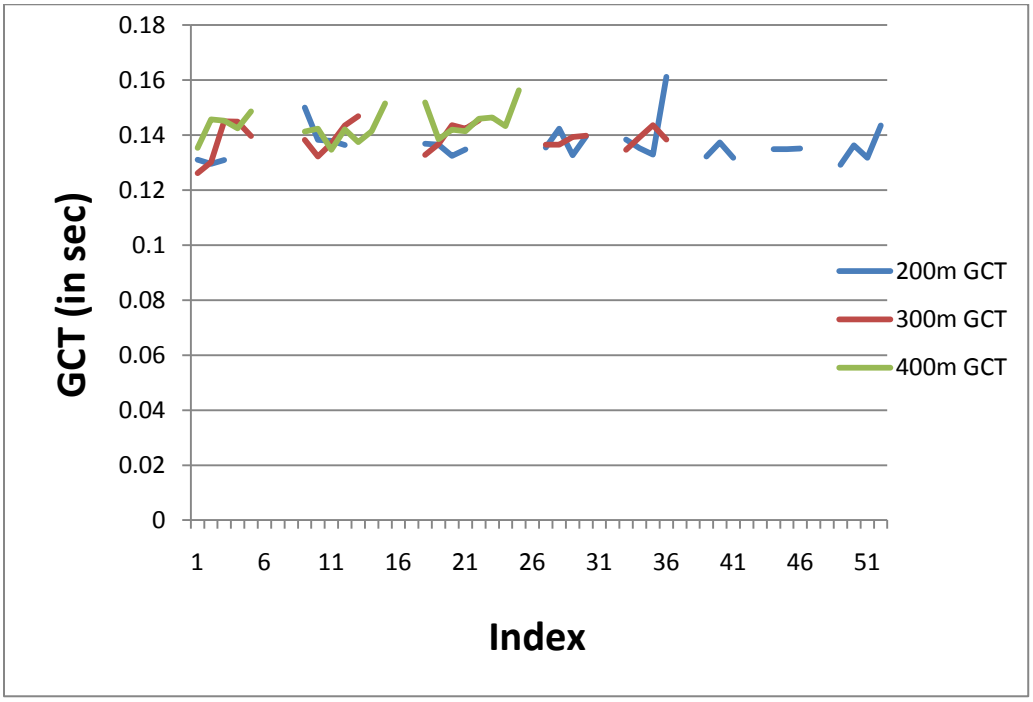


Figure 5.41: Progression of GCT_S for the different runs by Anna

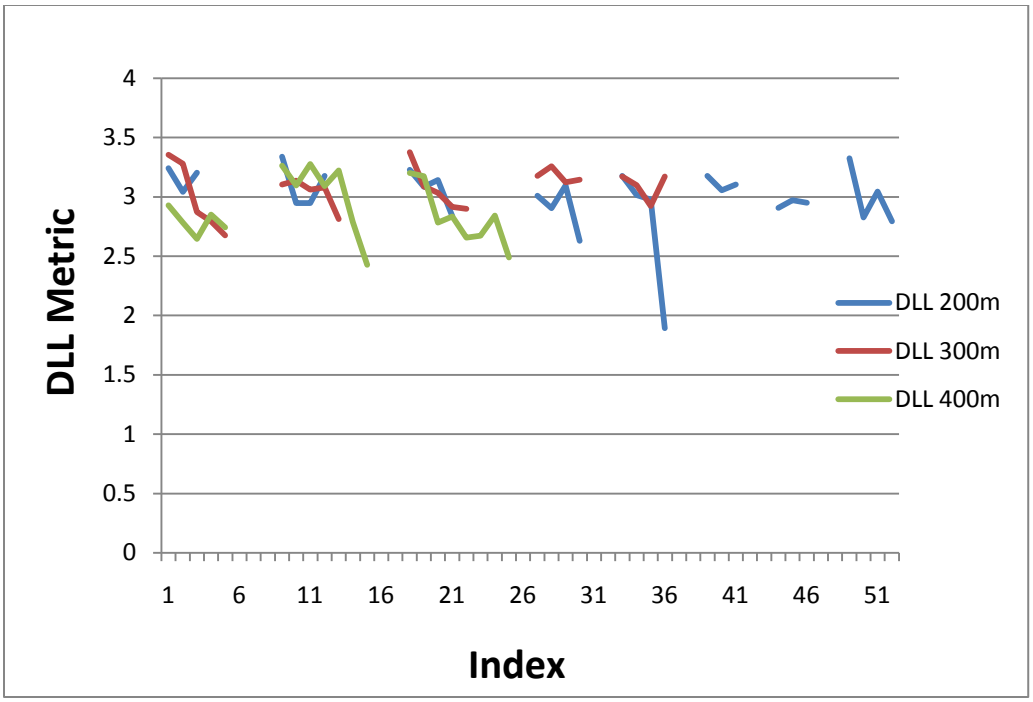


Figure 5.42: Progression of DLL metric for the different runs by Anna

The trends observed here are much the same, thus consolidating the findings. 200m run metric values do not show a consistent trend as can be expected for such a short run. Metric values for 300m and 400m runs follow the expected trend. Based on the results of these three runners, some of the observations are summarized in the section below.

5.5.4 Key Observations

1. The rate of drop in performance over these short runs, as compared to the marathons, has been much higher. Clearly running pace is the driving factor. Faster the run, higher is the rate of lactic acid build up in the muscles and easier it is to get tired. The findings here are an important first step to pin pointing the exact relationship between pace and metric values
2. Angela's performance, while still being in the elite range, was worse as compared to other runners. On inquiry it was found that she recently joined the OTC and

Terrence Mahon, a world class coach confirmed that her technique needed improvements. Terrence was excited to see the metric development and the system's ability to give pointers.

6 CONLUSIONS AND FUTURE WORK

Using the advancements in the sensor technology and mobile computing facilities, we can now design real time human performance evaluation systems. We have restricted ourselves to characterizing the running performance of elite middle distance runners here. It was our endeavor to deconstruct the biomechanical running cycle into a set of objectively measurable metrics that provide a natural intuition into the running technique. We designed a set of monotonic, physics based, computationally efficient real time metrics which break down each running stride into easily comprehensible objective values.

Tri-axial accelerometer data was used to develop six monotonic metrics. We developed an event segregation model for isolating running data and a robust stride recognition algorithm to segment strides. Cadence, GCT and DLL turned out to be highly proficient and insightful metrics with each contributing towards the overall efficiency factor. The overall efficiency factor was able to accurately model the runner performance over the course of the race. CDEL, Over-stride and Pronation also provided key information about the running technique of a runner. Product testing was done in rigorous race conditions and data sets from several elite runners were analyzed in real time to consolidate and prove the findings presented in this thesis. Several key recommendations to the runners were made based on the system findings.

From the data analysis of some of the elite runners, many key conclusions can be drawn. Firstly, the effect of fatigue is clear. As the race progresses, the overall efficiency has a tendency to drop for most runners. This is because of a drop in the distal leg lift and cadence and an increase in the ground contact time with distance. These relationships bear

high importance for runners, as now using this system personalized training modules can be developed for each runner. Endurance limits can now easily be tracked and injury prevention measures can be taken to correct technical flaws using our analysis system.

Special attention needs to be given towards the development of statistical models which group runners based on their basic physical features like height and weight. We have seen that even though all the metrics are monotonic, some of them need to be properly normalized. While it is clear that metrics like DLL and CDEL are a function of height and weight of the runner, their specific relationships are still a mystery. Having data from several runners would help build a model to counter this problem. Work is being done in this direction and we are constantly trying to leverage any relationships in the running community to get data from elite runners.

Another interesting problem that this work pointed out is the role of gravity in the accelerometer data. As the running stride varies in magnitude and style (with runner and with pace), we cannot rely on a predetermined gravity profile. This brings to the forefront the challenge of creating a gravity profile on the fly. This can then be used to get only foot acceleration data which will increase our understanding of the gait cycle even more. It is a non-trivial problem which requires use of other sensors and much research and development.

It is also worth mentioning here that there are many opinions in the world of professional running as to what constitutes a perfect stride. This work, by no means, tries to address that debate. We simply endeavor to recognize physical features through objective metrics. This is one of the reasons why the overall efficiency metric is based on only a few basic metrics. It is my belief that as we develop our understanding of the gait

cycle through the system of metric performance evaluation, many of the existing questions will answer themselves.

This work provides an excellent platform for human performance evaluation techniques and sets a benchmark for similar products to follow. Much remains desired in terms of human performance evaluation in health recovery programs and sports. People want to see a set of objective metrics through which they can compare themselves to elite performers in their field of interest. To achieve these objectives we need to collect more data, globalize all metrics through intelligent normalization and come up with even more insightful evaluation methods. We believe that the field of Wireless Health Monitoring (WHM) is still in its nascent stages and several exciting opportunities and challenges lie ahead.

Bibliography

[1] R. Requa, L. DeAvilla and J. Garrick, “Injuries in recreational adult fitness activities,” *The* American *Journal of Sports Medicine*, vol. 21, no. 3, p. 461, 1993.

[2] Keith Williams and Peter Cavanagh, "Relationship between distance running mechanics, running economy, and performance," *American Physiological Society*, pp. 1236-1245, 1987.

[3] P. Fleming, C. Young, S. Dixon, and M. Carre, “Athlete and coach perceptions of technology needs for evaluating running performance,” *Sports Eng.*, vol. 13, pp. 1–18, 2010.

[4] Tom Novacheck, "The biomechanics of running," *Gait and Posture*, pp. 77-95, 1997.

[5] M. Estivalet, P. Brisson, M. Wijnen, M. B. Hoppenbrouwers, and J. W. M. Willems, “Runalyser: Real Time Analysis of Running Technique in Practice,” *The Engineering of Sport 7. Springer,* 2008. [Online]. Available: http://dx.doi.org/10.1007/978-2-287-99056-4n 3

[6] Nike. (2011) Nike+ Sportband User Guide. [Online]. Available: http://nikeplus.nike.com/nikeplus/utility/v1/en_US/help.html

[7] Polar USA. (2011) RS100 Training Computer. [Online]. Available: http://ww.polarusa.com/

[8] J. Abt, T. Sell, Y. Chu, M. Lovalekar, R. Burdett and S. Lephart, “Running kinematics and shock absorption do not change after brief exhaustive running,” *The Journal of Strength & Conditioning Research*, vol. 25, p. 1479, 2011.

[9] K. Moran, M. Clarke, F. Reilly, E. Wallace, D. Brabazon, and B. Marshall, "Does endurance fatigue increase the risk of injury when performing drop jumps?," *The Journal of Strength & Conditioning Research*, vol. 23, no. 5, p. 1448, 2009.

[10] J. Mizrahi, O. Verbitsky, E. Isakov, and D. Daily, "Effect of fatigue on leg kinematics and impact acceleration in long distance running," *Human Movement Science*, vol. 19, no. 2, pp. 139 – 151, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/B6V8T-410MHHK-1/2/75b3943bf9f731edd9ea54295ce89d1

[11] T. Burgess and M. Lambert, "The effects of training, muscle damage and fatigue on running economy: review article," *International SportMed Journal*, vol. 11, pp. 363 – 379, 2010.

[12] J. Daniels, "A physiologist's view of running economy," *Medicine and Science in Sports and Exercise*, vol. 17, no. 3, p. 332, 1985.

[13] J. Larson, "Design and Development of Performance Metrics for Elite Runners," *Master's thesis*, University of California, Los Angeles, 2011.

[14] G. Pottie and W. Kaiser, *Principles of Embedded Networked Systems Design*. New York: Cambridge University Press, 2005.

[15] T.K. Sethuramlingam and A. Vimalajuliet, "Design of MEMS Based Capacitive Accelerometer," in *International Conference on Mechanical and Electrical Technology*, pp. 565-568, 2010.

[16] J. Chae, H. Kulah, and K. Najafi, "A monolithic three-axis micro-g micromachined silicon capacitive accelerometer," *Journal of Microelectromechanical Systems*, vol. 14, no. 2, pp. 235 - 242, April 2005.

[17] Analog Devices, Inc., ADXL326 Datasheet, August 2009.

[18] A. Chandrakasan et al., "Design considerations for distributed microsensor systems," in *Proceedings of the IEEE Custom Integrated Circuits*, pp. 279-286, 1999.

[19] Motorola, Inc. (2011, September) DROID BIONIC by Motorola fact Sheet. [Online]. http://mediacenter.motorola.com/Fact-Sheets/DROID-BIONIC-by-Motorola-Fact-Sheet-353a.aspx

[20] NXP Semiconductors, LPC2141/42/44/46/48 Product Data Sheet, August 12, 2011.

[21] Sparkfun Electronics, Logomatic v2 Datassheet, January 9, 2009.

[22] R. Blickhan, "The spring-mass model for running and hopping," *Journal of Biomechanics*, vol. 22, no. 11-12, pp. 1217–1227, 1989.

[23] M. Eriksson, K. Halvorsen, and L. Gullstrand, "Immediate effect of visual and auditory feedback to control the running mechanics of well-trained athletes," *Journal of Sports Sciences,* vol. 29, no. 3, pp. 253–262, 2011.

[24] P.G. Weyand, D.B. Sternlight, M.J. Bellizzi, S. Wright, "Faster top running speeds are achieved with greater ground forces not more rapid leg movements," *Journal of Applied Physiology* **89**, pp. 1991–1999, 2000.

[25] Roger A. Mann, Gary T. Moran and Sandra E. Dougherty, "Comparative electromyography of the lower extremity in jogging, running, and sprinting", *American Journal of Sports Medicine* **14**, pp. 501-510, 1986.

[26] T.D Bushnell, I Hunter, *Sports Biomechanics* **6**, pp. 261-268, 2007.

[27] D.E. Lieberman et al., "Foot strike patterns and collision forces in habitually barefoot versus shod runners," *Nature* **463**, pp. 531–535, 2010.

[28] A. Nummela, T. Keranen, and L. Mikkelsson, "Factors related to top running speed and economy," *International journal of sports medicine*, vol. 28, no. 8, pp. 655–661, 2007.

[29] T. Anderson, "Biomechanics and running economy" *Sports medicine (Auckland, NZ)*, vol. 22, no. 2, p. 76, 1996.

[30] O. Storen, J. Helgerud, & J. Hoff, "Running stride peak forces inversely determine running economy in elite runners," *Journal of Strength & Conditioning Research*, Jan 21 2010.

[31] H. Kyrolainen, A. Belli, and P. Komi, "Biomechanical factors affecting running economy," *Medicine & Science in Sports & Exercise*, vol. 33, no. 8, p. 1330, 2001.