PREDICTING THE LEVEL OF KNEE OSTEOARTHRITIS BY USING DIMENSION REDUCTION TECHNIQUES ON TIME SERIES FEATURES OF GAIT DATA

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SALİH CANER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARMENT OF
HEALTH INFORMATICS

DECEMBER 2013

PREDICTING THE LEVEL OF KNEE OSTEOARTHRITIS BY USING DIMENSION REDUCTION TECHNIQUES ON TIME SERIES FEATURES OF GAIT DATA

Submitted by **SALİH CANER** in partial fulfillment of the requirements for the degree of **Master of Science in Health Informatics, Middle East Technical University** by,

Prof.Dr. Nazife Baykal
Director, **Informatics Institute, METU** _______________

Assist.Prof.Dr. Yeşim Aydın Son _______________
Head of Department, **Health Informatics, METU**

Prof.Dr. Ünal Erkan Mumcuoğlu _______________
Supervisor, **Health Informatics**, **METU**

Dr. Nigar Şen _______________
Co-Supervisor, **Space and Defence Technologies**

Examining Committee Members

Assist.Prof.Dr. Yeşim Aydın Son _______________
Health Informatics, METU

Prof.Dr. Ünal Erkan Mumcuoğlu _______________
Health Informatics, METU

Dr. Nigar Şen _______________
Space and Defence Technologies

Assist.Prof.Dr. Didem Gökçay _______________
Health Informatics, METU

Assist.Prof.Dr. Tolga Özkurt _______________
Health Informatics, METU

Date: 02 December 2013

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Salih Caner
Signature :

ABSTRACT

PREDICTING THE LEVEL OF KNEE OSTEOARTHRITIS BY USING DIMENSION REDUCTION TECHNIQUES ON TIME SERIES FEATURES OF GAIT DATA

Caner, Salih
Department of Health Informatics
Supervisor: Prof.Dr. Ünal Erkan Mumcuoğlu
Co-Supervisor: Dr. Nigar Şen

December 2013, 98 pages

Gait analysis is the systematic study of human walking. Gait analysis is also used on facilitating medical diagnosis decisions such as on a disease called knee osteoarthritis (OA) which is a disorder that affects joint cartilage. Gait data used in this study was collected from Ankara University Department of Physical Medicine and Rehabilitation Gait Laboratory for the study called “A Decision Support System for Grading Knee Osteoarthritis Using Gait Data” implemented by Şen. Some of the main obstacles of analyzing gait data which was captured by special video cameras are multidimensionality and correlation of data. Due to the complexity and high dimensionality of gait patterns, clinicians have limited understanding and difficulty interpreting raw data. The objective of this study is to implement some dimension reduction techniques on the waveforms of gait data to improve accuracy of classification methods in grading OA.

Keywords: Gait analysis, knee osteoarthritis, dimension reduction, classification

ÖZ

YÜRÜYÜŞ VERİLERİNİN ZAMAN TABANLI PARAMETRELERİNİN BOYUT KÜÇÜLTME YÖNTEMLERİ KULLANILARAK DİZ OSTEOARTRİTİSİ SEVİYESİNİN TAHMİN EDİLMESİ

Caner, Salih
Master, Sağlık Bilişimi
Tez Yöneticisi: Prof.Dr. Ünal Erkan Mumcuoğlu
Ortak Tez Yöneticisi: Dr. Nigar Şen

Aralık 2013, 98 sayfa

Yürüme analizi insan yürüyüşünün sistematik olarak incelenmesidir. Yürüyüş analizi diz osteoarthritisi adı verilen eklem bölgesindeki bozuklukla ilgili hastalığın tanısı için de kullanılmaktadır. Bu çalışmada kullanılan veriler Şen tarafından gerçekleştirilen “A Decision Support System for Grading Knee Osteoarthritis Using Gait Data” isimli çalışma için Ankara Üniversitesi Fizik Tedavi ve Rehabilitasyon Merkezi’nde toplanmıştır. Özel kameralarla elde edilen yürüyüş verisinin analiz edilmesindeki en büyük sorun dalga biçimli verilerin yüksek miktardaki boyutu ve parametrelerin bağımlı oluşudur. Yürüyüş verilerinin karmaşıklığı ve yüksek miktarda olmasından dolayı hekimlerin bu verileri anlaması ve yorumlama kabiliyeti sınırlı olmaktadır. Bu çalışmanın amacı yüksek boyutlu verilerin bazı boyut küçültme teknikleri ile küçültülerek veri sınıflandırma methodlarının hastalığın derecesini belirlemekteki başarısını artırmaktır.

Anahtar kelimeler: Yürüme analizi, diz ostearthritisi, boyut küçültme, sınıflandırma

To My Family

ACKNOWLEDGEMENTS

I would like to express my sincere appreciations to Prof.Dr. Ünal Erkan Mumcuoğlu for his guidance and insight throughout the research. I would like to thank Dr. Nigar Şen for her valuable and constructive suggestions and sharing her possessions concerning my thesis. I am also grateful to my committee members Assist. Prof.Dr. Didem Gökçay, Assist. Prof.Dr. Yeşim Aydın Son and Assist. Prof.Dr. Tolga Özkurt for their constructive comments. I would like to express my sincere appreciation to my dear wife Sema and to my dear sons Ali Kaan, Eren and Erdem for their patience and sacrifice throughout my study. I would like to thank my managers and colleagues at the Ministry of Foreign Affairs for their toleration to complete my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASIS Anterior Superior Iliac Spine
BMI Body Mass Index
CA Correspondence Analysis
DFT Discrete Fourier Transform
EMG Electromyography
FA Factor Analysis
FFT Fast Fourier Transform
GRF Ground Reaction Forces
HD Hausedorff Distance
KPCA Kernel Principal Component Analysis
LDA Linear Discriminant Analysis
MD Mahalanobis Distance
NN Neural Network
OA Osteoarthritis
PC Principal Component
PCA Principal Component Analysis
PDM Point Distribution Model
SFS Sequential Forward Selection
SVM Support Vector Machine
APABD Ankle Abduction Power
APDOR Ankle Dorsiflexion Power
APROT Ankle Rotation Power
APTOT Total Ankle Power
FDOR Foot Dorsiflexion
FMABD Foot Abduction Moment
FMDOR Foot Dorsiflexion Moment
FMROT Foot Rotation Moment
FPRO Foot Rotation Power
FROT Foot Rotation
HABD Hip Abduction
HFLEX Hip Flexion
HMABD Hip Abduction Moment
HMFLEX Hip Moment Flexion
HMROT Hip Moment Rotation
HPABD Hip Power Abduction
HPFLEX Hip Power Flexion
HPROT Hip Power Rotation
HPTOT Hip Power Total

HROT Hip Rotation
KFLEX Knee Flexion
KMFLEX Knee Flexion Moment
KMROT Knee Rotation Moment
KMVAL Knee Valgus Moment
KPFLEX Knee Flexion Power
KPROT Knee Rotation Power
KPTOT Total Knee Power
KPVAL Knee Valgus Power
KROT Knee Rotation
KVAL Knee Valgus
POBLIQ Pelvis Obliquity
PROT Pelvis Rotation
PTILT Pelvis Tilt

CHAPTER I

INTRODUCTION

1.1 Motivation

A large number of diseases affect the neuromuscular and musculoskeletal systems and may thus lead to disorders of human walking (gait) . Decreasing the cost and increasing the ability of specifications of gait analysis systems make clinicians prefer these systems to compare pre and post operative gait patterns and to distinguish between primary problems and coping responses. On the other hand, because of high dimensionality and correlation of kinematic, kinetic, electromyography (EMG) parameters of time-dependent waveform of gait data, analyzing gait data is a complicated task. Additionally gait recordings change trial to trial and session to session. As researchers have been studying new ways to overcome these barriers, it has been accepted that there is no sufficient and reliable technique which keeps useful information from correlated and high dimensional gait variables on reducing gait data [1]. Due to the complexity and great amount of waveformed data, understanding and distinguishing gait data is limited by the clinicians and knowledge about the selection of suitable analysis technique in diagnosing different gait abnormalities is also limited [2].

Complex and high dimensional gait data produced by the gait analysis systems needs smart solutions to help the clinicians to manage and extract useful information.

1.2 Scope

Because of the challenges of analyzing the temporal waveforms representing joint angles and power measures of the gait data, there are numerous studies to help to analyze and extract the meaningful information of gait data. The most commonly method used analyzing the temporal gait data is to pick discrete instants such as peak values or ranges. However, extracting special points is subjective and does not take into account of the whole data which might hold valuable information. The standing obstacle analyzing the gait data is effectively reduction of the dimension as most pattern recognition techniques suffer from classifying high dimensional data (curse of dimensionality). That is why various dimension reduction methods have to be investigated to better summarize huge amount of gait data. The scope of this study was to use as many gait features as possible while reducing effects of curse of dimensionality on classifications techniques.

1.3 Objective

The objective of this study was to examine the contribution of dimension reduction techniques on the prediction of the grade of knee osteoarthritis (OA) by using the kinetic and kinematic features of gait data. As 3D motion analysis systems produce huge amount of waveform data which holds valuable information, data reduction process is the key on accuracy of the classification. Therefore, the focus of this study was to reduce dimension of data while preserving as much information as possible.

1.4 Contributions

As the curse of dimensionality is the one of the main challenges in classification of gait data, the intensity of this study focused on the application of dimension reduction techniques on the waveform data. Different feature selection and feature extraction methods were applied to gait data. In previous work of this study, Mahalanobis Distance and Hausedorff distance metrics were used to select best features which might have best discriminating power classifying the grades of OA. Gait data was also reduced by signal downsampling methods according to analysis of frequency spectrum and application of low pass Butterworth filter to avoid aliasing which causes loss of information. Principal Component Analysis (PCA) and Kernel PCA are used as feature extraction methods. To simplify the classification process, the classification problem of gait data was split into a set of binary classification problems. Sequential Forward Selection (SFS) method was applied as the last step of dimension reduction process to select best attributes which make considerable contribution to classifiers. SVM was used as the classification method because it is accepted as successful classification method. After using two different classification paths, %83.75 and %76.25 overall success rates were achieved. It was observed from the results that Grade 1 had the lowest classification rate. Despite Kellgren & Lawrence (K&L) metric is accepted as the gold standard on grading knee osteoarthritis, it was also criticized in medical literature. Some medical reports also recommend not to take into account of the Grade 1 in gait analysis because of inconsistencies on that grade.

1.5 Organization of the Study

Motivation, objective, scope and contribution of this study were introduced in Chapter 1. Background information on osteoarthritis, gait and gait analysis were given in Chapter 2. In Chapter 3, information about the gait data used in this study was presented by defining collection procedures, specifications and by explaining interpretation of gait data. In Literature Review section of Chapter 4, challenges of analyzing gait data, the importance of dimension reduction regarding the analyzing gait data and academic studies and methods on analyzing gait data were explained. Additionally, explanation about the dimension reduction techniques in general and the methods used in this study were presented. Implementation of these methods were realized by executing two scenarios using combination different techniques and gait data sets. The result of these experiments was presented in Chapter 5. Finally, the discussions of the results and future works were explained in the Conclusion Chapter.

CHAPTER II

BACKGROUND INFORMATION ON OSTEOARTHRITIS AND GAIT ANALYSIS

2.1 Gait

The word “gait” describes ‘the manner or style of walking’, rather than the walking process itself. It thus makes more sense to talk about a difference in gait between two individuals than about a difference in walking [3]. Human gait is a complex process as the locomotors system incorporates input from the cerebellum, the motor cortex, and the basal ganglia, as well as feedback from visual, vestibular, and proprioceptive sensors [4].

2.2 Clinical Gait Analysis

Gait is the result of coordination and integration of different body systems such as central and peripheral nervous system (locomotor generator), muscles and skeletal levers, visual, proprioceptive, cognitive and cardiovascular systems. Any impairment in these body structures and functions may cause a pathological gait pattern. Individuals compensate the problem by changing their gait pattern so the resultant gait pattern is the combination of primary deficit and compensatory substitutions including but not limited to inadequate, excessive, inappropriately timed, or out-of-phase muscle action [5]. Diagnosis and the treatment of the abnormal gait are commonly decided based on clinical examination and observational gait analysis. Observational gait analysis may be inadequate because various pathologies may cause similar gait patterns [6].

Visual gait analysis is the most complicated and versatile form of analysis available [3]. However, observational gait analysis may not be sufficient because of the complex nature of the gait pattern, the limitations of the human observation and subjectivity of the assessment due to requirement of the high expertise. Additionally, gait force parameters cannot be observed visually. In a study on the reproducibility of visual gait analysis, Krebs et al. (1985) found it to be ‘only moderately reliable’.

Although videotaping gait and analyzing it in slow motion may help to make better assessment by overcoming human perception, it is not an objective method since it does not provide quantitative data in the form of numbers. However, it provides reassessing and reanalyzing the gait after a period of time [3]. The more reliable examination can be implemented by the using three-dimensional (3D) quantitative gait analysis. The long range of parameters of 3D quantitative gait analysis has helped clinicians to diagnose the problem more accurately than the observational analysis [7].

2.3 Three-dimensional Quantitative Gait Analysis

Three-dimensional quantitative gait analysis systems have three main outputs which are kinematic, kinetic, and dynamic electromyographic measurements. Typically, gait data is collected approximately 1-2 hours and assessment of the data takes 1-2 hours for each patient. The main requirements for patients are to have a height of at least 1 meter, sufficient cognitive ability to follow the instructions and ability to walk 5 or 10 times at the walking path. Subject preparation, data recording and data analysis are the respective steps of the analysis. Firstly, basic parameters including height, weight, leg length and joint width of the knee and ankle are collected. Secondly, special markers are placed on specific anatomical places such as knee, hip, tibia, toe, ankle and anterior superios iliac as seen in Figure 2.1.

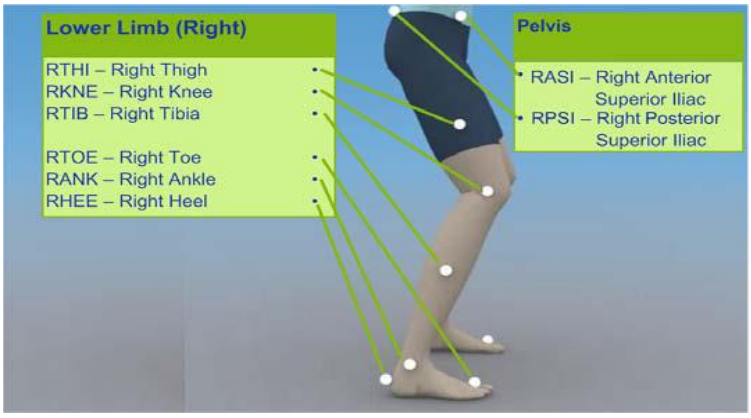


Figure 2.1: Standard lower body marker placement (courtesy to Vicon [58])

When special markers are placed to the patients, they are requested to walk at a self-selected pace for a number of times over the walking path. Cameras record the three dimensional location of each marker as the subject walks. Ground reaction forces collected from the force plates and kinematic data combined with inverse dynamics are used to evaluate joint moments and powers of the hip, knee, and ankle joints in spatial locations. Afterwards, the recorded data can be examined by the medical experts [6].

2.4 Knee Osteoarthritis

In this study, gait data was used to classify the level of knee OA which is a chronic disease of the joint cartilage and bone.

Movement of the body parts and flexibility of various parts of the body are achieved by joints and muscles which are attached to the bone. Synovial fluid which is small amount of thick fluid located between the cartilage which is a hard, smooth tissue covering the end of bones, lubricates the joint so that smooth movement between the bones can be provided.

Joint cartilage deforms and the tissue around the cartilage can also be damaged at the joints of a patient having OA and affected area can be seen around the joint edges [8, 14] (Figure 2.2).

Pain and stiffness in the joints are the most common symptoms of OA. The damaged joint becomes larger than normal since inflammation and deformation of the bone around the cartilage. OA is usually observed on most used joints such as hips, knees, finger joints, thumb joints and lower spine [9].

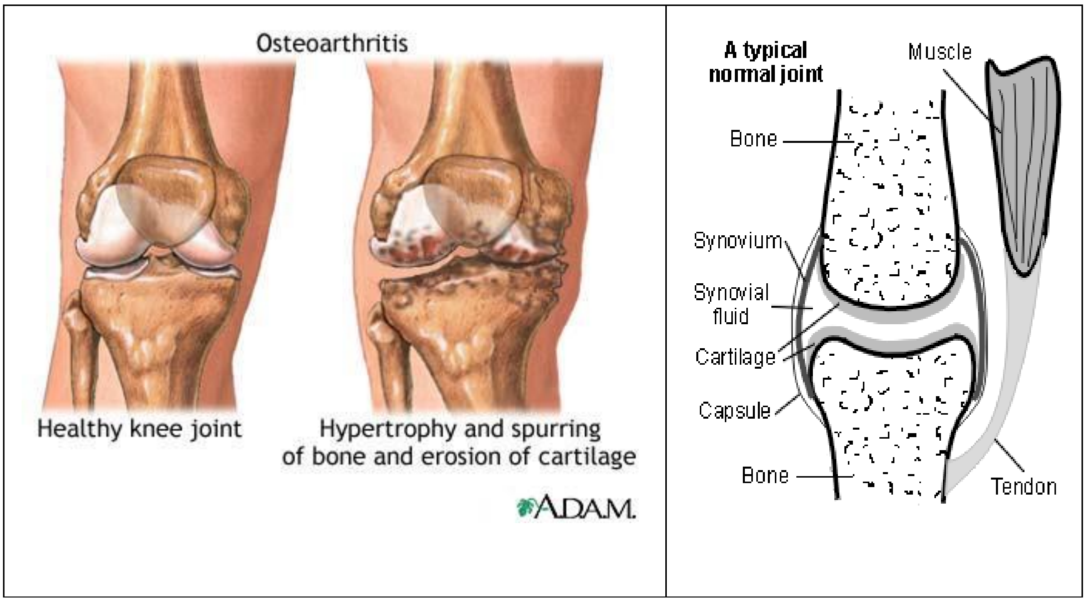


Figure 2.2: Left: Normal Joint and joint with OA. Right: Joint cartilage and the bone tissue (courtesy to [14, 62])

OA is the most prevalent in older adults and knee is the second most commonly afflicted joint of OA [10]. It is estimated that 9% of men and 18% of women over age 65 have knee OA [8]. Individuals with knee OA experience pain, stiffness, and decreased range of motion of the joints. These symptoms significantly limit an individual's ability to perform daily activities and may cause to loss of functional independence [11].

Diagnosis of the disease is made according to symptoms, physical examination and XR images. XR images show evidence of OA especially in joints enduring weight of the body like hip and knee. Kellgren-Lawrence method is used for radiological assessment of OA [12].

Kellgren-Lawrance method defines five levels of OA which are explained below.

Level 0, indicates a definite absence of x-ray changes of OA.
Level 1, indicates doubtful narrowing of joint space and possible outgrowth of the bone.
Level 2, indicates definite growth of the bone and possible narrowing of joint space.
Level 3, moderate multiple outgrowths, definite narrowing of joint space, some sclerosis and possible deformity of bone contour.
Level 4, large outgrowths, marked narrowing of joint space, severe sclerosis and definite deformity of bone contour.

CHAPTER III

GAIT DATA

3.1 Data Collection

The gait data used in this study was primarily collected in Ankara University Faculty of Medicine, Department of Physical Medicine and Rehabilitation Gait Laboratory. The gait data set was used for the PhD study called "'OAGAIT': A Decision Support System for Grading Knee Osteoarthritis Using Gait Data" implemented by Şen [13].

Agreements of subjects for the usage of data were received before gait analysis. Subjects underwent the same procedure and examined and instructed by the same physician. Kinetic and kinematic features of gait was collected in the laboratory using computer-interfaced video cameras which detect markers attached to the body of the subject and using force platforms which calculates the forces and moments produced between the patients and the ground. Data acquisition was implemented by Vicon 370 Motion Measurement and Analysis System which has 5 special video cameras to track the markers and a computer system for data processing of the captured data.

Retro-reflective markers were placed on special locations of joints to prevent artifacts because of skin movement and on specific anatomical of locations of the lower body such as anterior superior iliac spine (ASIS), sacrum, lateral thigh, joint line of the knee. Ground reaction forces (GRF) were collected using 2 force plates. Moments and ground reaction forces were normalized with body weight and height for the purpose of comparison with other results found in the literature. Position of reflective markers was sampled 60 times a second [6].

3.2 Data Characteristics

Total of 133 normal and abnormal subject's statistics are given in Table 3.1. As it can be seen in Table 3.1, the number of abnormal female subjects is greater than the male subjects. Despite the gait data has been collected from one laboratory and there has been limited number of subjects, the difference between number of male and female subjects can be interpreted as that OA is mostly seen on females. Because of left and right leg waveforms were recorded separately, 180 records of 133 subjects were used in this study.

Table 3.1: General characteristics of gait data

	Number of Subjects	Male	Female	Mean (Age)	Standard Deviation (Age)	Mean (BMI)	Mean (Ache Level)
Level 0	34	15	19	43.1	15.2	27.3	1
Level 1	18	4	14	58.4	7.8	45.9	5.7
Level 2	48	5	43	59	8.2	32.1	6.8
Level 3	33	4	29	63	7.9	33.1	7.3

3.3 Time-distance Features

The well-known gait parameters providing the basic form of objective gait evaluation (also known as the temporal and distance factors of gait) are stride length, cycle time (or cadence), and speed [3].

Time-distance parameters were gathered at the end of one gait cycle. Vicon Motion Analysis System provides the following parameters:

- **Cadence** measures number of steps in a given period.
- **Walking speed** is the distance covered in meters per second.
- **Stride time** is the duration of gait cycle.
- **Step time** is the duration between touching ground with left and right foot.
- **Single support** is the duration of time when one foot is in contact with the ground during the gait cycle.
- **Double support** is the duration of time when both feet are on the ground during the gait cycle.
- **Stride length** is the distance of one gait cycle.
- **Step length** is the distance between the right and left of length of a complete gait cycle.

Analyzing of after treatment effects using time-distance measurements enables useful information about the patient’s walking ability and reinforces the clinical evaluation.
However, time-distance measurements are only end products of a complicated motion pattern that neither explain a gait pattern nor distinguish between the primary gait fault and compensation [7]. On the other hand, Şen made predictions on discriminating levels of OA by these features. A decision tree was constructed by combining subject’s personal features such as age, body mass index (BMI), pain, stiffness, period, history and sex with the time-distance parameters. On the leaf of the decision tree a neural network is used to differentiate the level of the knee osteoarthritis. As a result, overall % 80 success rate has been achieved by combining decision tree and neural network outputs.

3.4 Personal Features

Body Mass Index (BMI), age, ache level and first move ache level were used as personal features.

3.5 Kinetic and Kinematic Features

The kinetic and kinematic features of the gait are gathered by 3D analysis of lower section of the body which is described by using the following reference planes:

Sagittal plane is the vertical plane which divides the body into right and left sections.
Frontal plane separates the body into two sections as posterior and anterior.
Transverse plane is the horizontal plane which divides the body into interior and superior sections (Figure 3.1).

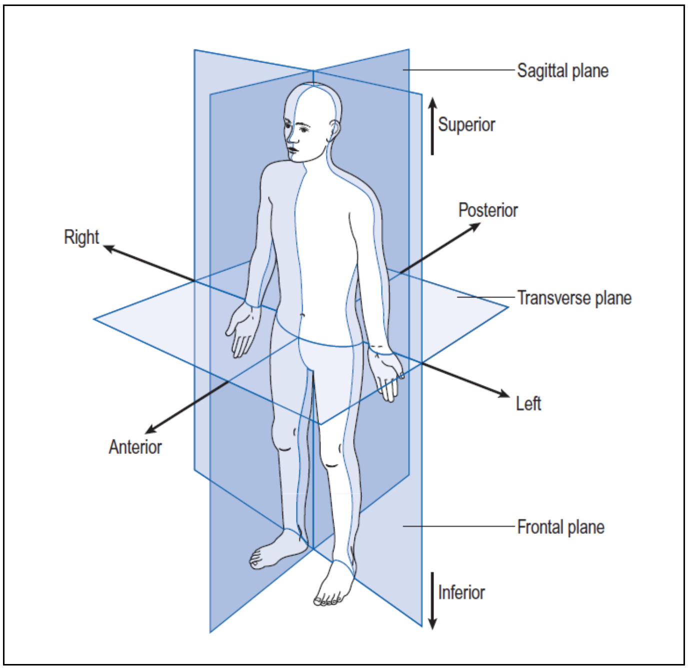


Figure 3.1: Body reference planes (courtesy to Whittle [3])

3.6 Terms Used in Gait Analysis

The gait features used in this study were named depending on direction of the movement according to the planes and section of the lower body (Figure 3.2).

The movement occurred in the sagittal plane at the hip and knee is called flexion and extension while the movement of the ankle in the same plane is called dorsiflexion and plantar flexion. The motion in the frontal plane is called abduction and adduction and movement taking place in the transverse plane is called internal and external rotations. Angulations of a joint towards or away from the midline are called varus and valgus, respectively; knock knees are in valgus, bow legs are in varus. D*istal* means away from the rest of the body: the fingers are the distal part of the hand.

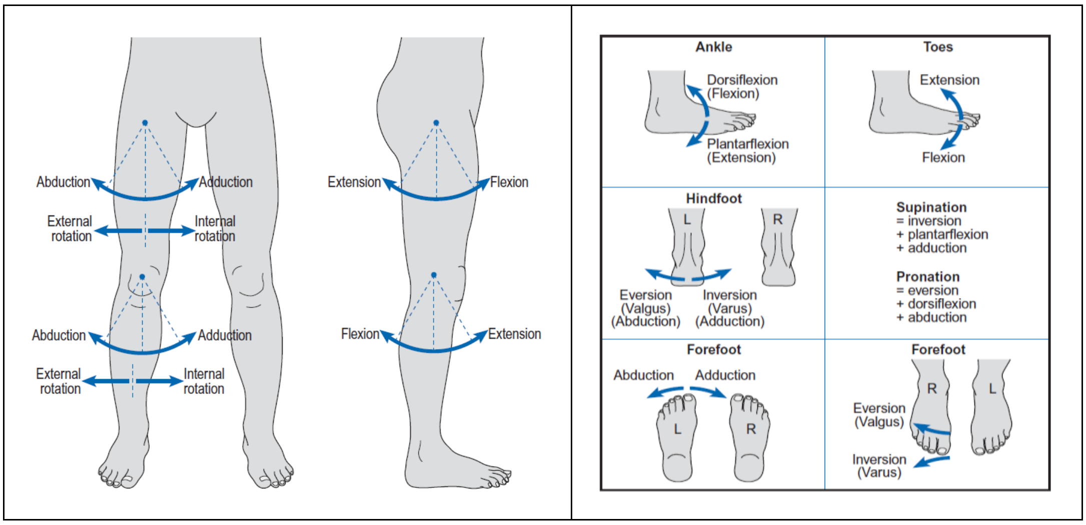


Figure 3.2: Left: Movements about the hip joint and knee joint on sagittal and frontal plane. Right: Movements of the ankle, toes, hind foot and forefoot on sagittal and frontal plane. (Courtesy to Whittle [3])

Most joints can only move in one or two of these three planes mentioned above. Most of the body sections take part in some way when we walk. However, from a practical point of view, pelvis and the legs have the important roles body sections for walking [3].

The names and the abbreviations of sample kinetic and kinematic features used in the study and their corresponding planes are given in Table 3.2.

Table 3.2: Name of kinetic and kinematic parameters of gait and their corresponding planes

Feature Name	**Marker's position**	**Plane+Moment/Power**
Pelvic Tilt (PTilt)	Pelvis	Sagittal
Pelvic Oblique (POblique)	Pelvis	Frontal
Pelvic Rotation (Prot)	Pelvis	Traverse
Hip Flexion Moment (HMFlex)	Hip	Sagittal+Force Plate
Hip Abduction Moment (HMAbd)	Hip	Frontal+Force Plate
Hip Rotation Moment (HMRot)	Hip	Traverse+Force Plate

3.7 Kinematics

"The *gait cycle* is defined as the time interval between two successive occurrences of the repetitive events of walking" [3]. Initial contact of a foot is mainly accepted as the start of the gait cycle and end of the cycle defined by recontacting the ground of the same foot. Gait cycle consists of two phases. First phase is called stance phase which starts with touching the right foot to the ground and lasts about 60% of gait cycle. The second phase is called swing phase which starts with taking off the right foot from the ground and lasts about 40% of gait cycle. On the other hand, these percentages vary as the speed of the walking changes. Stance phase has four sub phases and swing phase has three sub phases as seen in Figure 3.3. Kinematic data is the result of the combination of movement of joints in three dimensions, velocity and accelerations.

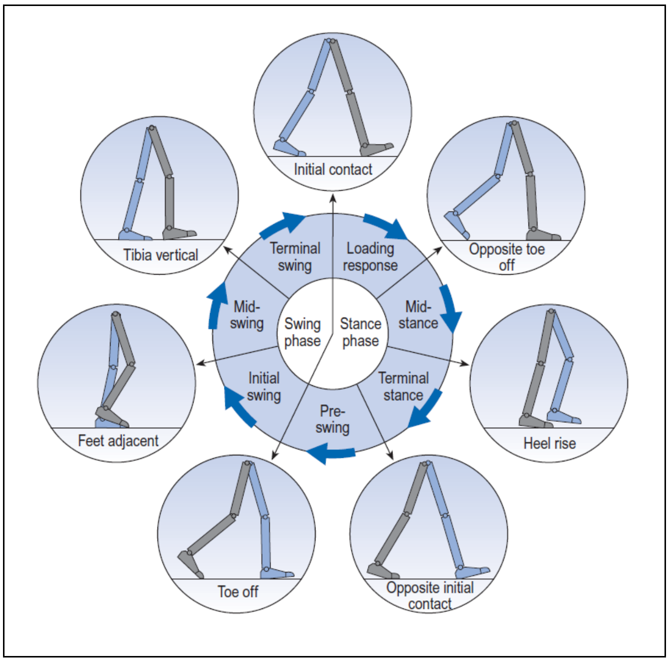
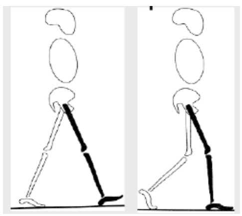
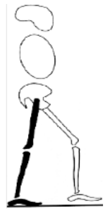
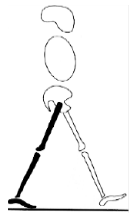


Figure 3.3: Position of the legs during phases of a gait cycle (Courtesy to Whittle [3])

The angle of the markers changes by time as the subject walks. For example, on Loading Response Phase hip remains on 25^{o} on sagittal plane while markers around knees flexes 0 to 15^{o} and the angle of ankle changes 0 to 15^{o}. The angle changes of the hip, knee and ankle on the sagittal plane for each phase of gait cycle are illustrated in Figure 3.4.

Gait cycle is captured by the sophisticated video camera systems known as optoelectronic system which monitors the markers in 3D. In kinematic graphs, range, baseline shift, timings of peak and bottoms, and pattern of the line are evaluated, whereas for kinetic graphs, main focus is on timing of peak and depth and pattern of the line [6].

Loading Response Phase	Midstance Phase	Terminal Stance Phase
 **Hip**: 25^{o} flexion **Knee**: 0^{o} - 15^{o} flexion **Ankle**: 0^{o} - 10^{o} plantar flexion	 **Hip**: 25^{o} flexion - 0^{o} **Knee**: 15^{o} flexion - 0^{o} **Ankle**: 10^{o} plantar flexion -> 5^{o} dorsiflexion	 **Hip**: 0^{o} - 20^{o} extension **Knee**: 0^{o} **Ankle**: 5^{o} dorsiflexion - 10^{o} dorsiflexion

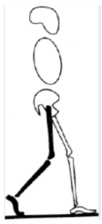
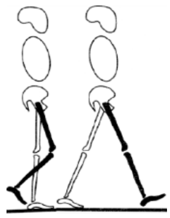
Preswing Phase	Swing Phase
 **Hip**: 20^{o} extension - 0^{o} **Knee**: 0^{o} - 40^{o} flexion **Ankle**: 10^{o} dorsiflexion - 20^{o} plantarflexion	 **Hip**: 0^{o} - 15^{o} flexion **Knee**: 40^{o}/60^{o} flexion ->0 **Ankle**: 20^{o} plantarflexion - 0^{o}

Figure 3.4: The change of angle of hip, knee and ankle on the sagittal plane **(**Courtesy to Patton [15])

Kinematic features produced by the Vicon Motion Analysis System are given in Table 3.3 with section of the body and their corresponding planes. During the gait cycle, angles of the sections of both legs measured by location changes of the markers are plotted as time progress. Graphs in Figure 3.5 illustrate variation of the degree of the angles for a complete gait cycle.

Table 3.3: Kinematic features produced by Vicon Motion Analysis System.

	Plane		
Anatomical Part	Flexion (Sagittal)	Abduction (Frontal)	Rotation (Traversal)
Pelvic	x	x	x
Hip	x	x	x
Knee	x	x	x
Ankle	x	x	x

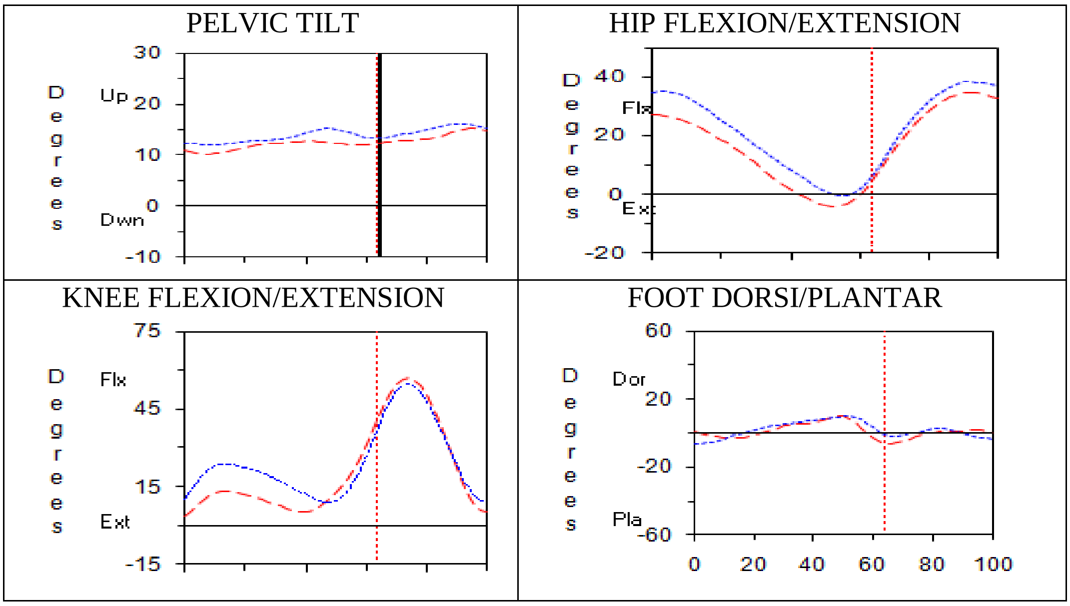


Figure 3.5: Deviation of angles of pelvis, hip, knee and ankle on sagittal plane during a single gait cycle for right (red) and left (blue) limbs (Courtesy to Vicon Motion Systems)

Generally speaking, the knee angle is defined as the angle between the femur and the tibia. The ankle angle is usually defined as the angle between the tibia and an arbitrary line in the foot. The ‘hip’ angle may be measured in two different ways: the angle between the vertical and the femur, and the angle between the pelvis and the femur [3].

3.8 Kinetics

Kinetics describes the joint moment and power calculated by GRF. GRF data is gathered from the force plates where the subject walks on. The vertical ground reaction forces change as the subject walks because of vertical upward and downward movement of the center of gravity.

The kinetic features are generated by the combination of GRF data, simplified models of the musculoskeletal system, inverse dynamics physics and kinematic information such as positions of related the markers, acceleration at each instant in time, angular velocity and moments of inertia.

An external joint moment (Nm/kg) refers to a net external load applied to the joint (ground reaction forces, segmental weight and inertia) and an internal joint moment is the result of the sum of all muscle activity acting about the joint in a given direction (forces from muscles, ligaments, and joint capsules). The joint moment tells us which muscles are acting at any given time. Mechanical power is derived by the combination of the joint moment and the joint angular velocity. The net power (=net joint moment x angular velocity W/kg) absorbed refers to eccentric muscular contraction, and net power generated refers to concentric muscular contraction. [7]

Names of the kinetic features used in this study and their corresponding anatomical parts are given in Table 3.4 and waveform graphs of kinetic features produced by the

Vicon Motion Analysis System are shown in Figure 3.6. As the interaction of the subject with the ground occurs at the stance phase, there is little activity at the swing phase because of limited or no contact with the ground.

Table 3.4 Kinetic features produced by Vicon Motion Analysis System.

	Joint Net Moments			**Joint Net Powers**			
Anatomical Part	Flexion	Abduction	Rotation	Total	Flex.	Abd.	Rot.
Hip	x	x	x	x	x	x	x
Knee	x	x	x	x	x	x	x
Ankle	x	x	x	x	x	x	x

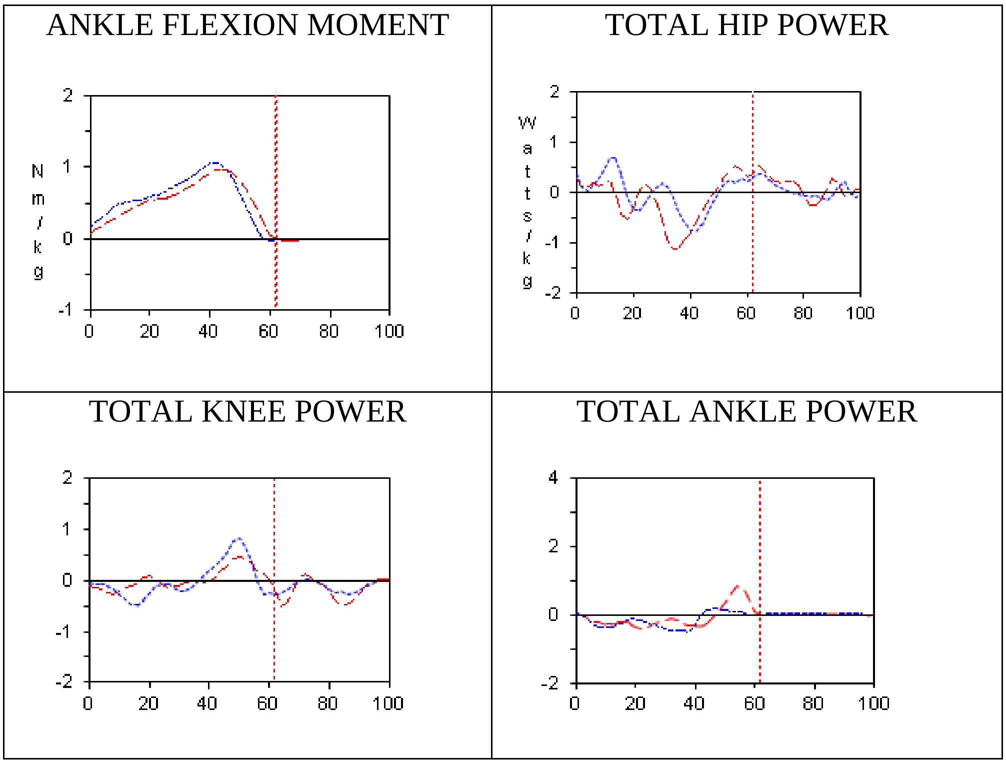


Figure 3.6 Deviation of moment and power hip, knee and ankle during a single gait cycle for right and left limbs (Courtesy to Vicon Motion Systems)

3.9 Interpretation of Gait Data

Interpretation of data is the most challenging part of gait analysis. Although 3D motion analysis system produces precise gait data and reflects the actual gait specification, the interpretation of generated data (as with many diagnostic procedures) varies expert to expert.

In addition to conventional examination techniques, well presented gait data and correct interpretation of data by the expert may help medical decision making.

3D gait analysis have some precision drawbacks because of the type and size of markers, different marker placement standards, length and age of the subject, stress, system errors, artifact and calibration errors as well as evaluator bias and poor training. There is a lack of standard to reduce errors so comparison of output the different laboratories is not possible. Evaluation of gait data may be improved by training the experts on the complexities of kinematic and kinetic and by the standardization of terminology [6]. Moreover, by applying pattern recognition techniques to expose valuable information on complex gait data, clinicians can better evaluate gait problems.

CHAPTER IV

DIMENSION REDUCTION

4.1 Backgroun Literature

4.1.1 Literature Review on Dimension Reduction

4.1.1.1 Dimension Reduction

Dimension reduction is transforming of data to a lower dimensional space such that uninformative variance in data is discarded, or such that a subspace in which data lives is detected [16].

To get better results on pattern recognition techniques, the number of samples of a data set should be significantly more than the number of features of the data set. It is desirable that number of samples of a data set should increase exponentially in accordance with the increase of the features of the data set if a relation needs to be defined between the samples and the features. On the other hand, in the most cases of real world pattern recognition problems, the number of samples of a data set is less than the number of features [17]. This phenomenon is one of the reasons of the poor performance of classification methods.

The curse of the dimensionality is first termed by Bellman in 1961 to express the fact that number of samples has to increase exponentially in accordance with the increase of the attributes of the sample to predict the function of the attribute to define the sample [18]. The performance of the classifier degrades as the dimension of data size increases or very few attributes exist to define the sample (Figure 4.1). It is general rule of thumb in classification that there should be number samples which are greater than ten times of features (d) ($n/d > 10$) [19]. Dimension reduction techniques are often applied as a data pre-processing step or as part of the data analysis to simplify the data model. The motivation of reducing the dimensionality can be summarized as follows.

- Data classification and clustering algorithms produce better results.
- Processing cost may be reduced.
- Irrelevant features can also reduce the prediction accuracy.

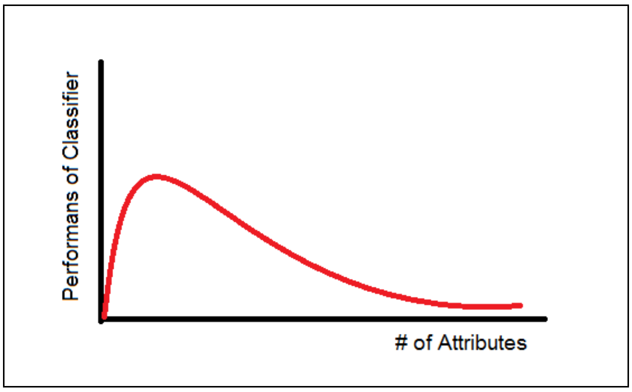


Figure 4.1: Curse of dimensionality

There are two methods for dimension reduction: feature selection and feature extraction. In feature selection, best k of d dimensions which gives the most information is sought and the other (d-k) dimensions are skipped. In feature extraction, a new set of k dimensions that are the combination of the original d dimensions are generated [20].

4.1.1.2 Feature Selection

(FS) algorithms seek the “best" minimum subset of the original features that give the most information about data (Figure 4.2). As the feature extractions methods transform the original data, transformed data set may not have physical meaning which is not a desired situation for knowledge discovery objectives. On the other hand, feature selection process keeps the original data so that data can be directly interpreted.

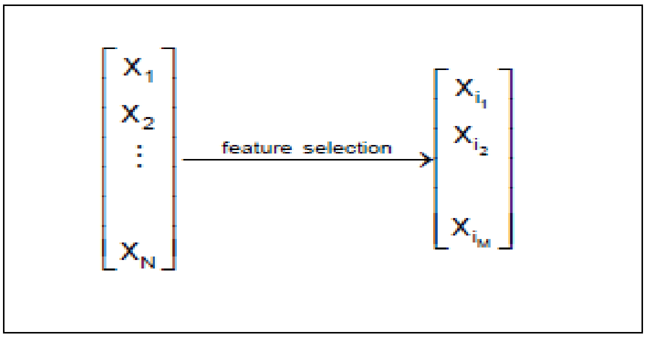


Figure 4.2: **Feature selection**: choosing a subset of all the features (the ones having more information about the characteristic of data)

In supervised learning, the objective of feature selection process is to pick the features that can predict the class label. In other words, the objective may be to select features that will construct the most accurate classifier. Feature selection for unsupervised learning (clustering) is a more complex issue and research into this field is recently getting more attention in several communities.

In supervised learning, selection process depends on a search strategy for exploring the space of feature subsets by searching a suitable starting point, determining successive candidate subsets and comparing the candidates. The evaluation schemes used in both supervised and unsupervised feature selection techniques can generally be divided into three broad categories.

Filter techniques evaluates the objective features by looking only at the intrinsic properties of data. Generally a score is calculated for each feature depending on the relevance and low-scoring features are removed. Finally, the resulting subset of features is presented as input to the classification algorithm [21].

Wrapper filter techniques seek the best subset by constructing the selection model depending classification algorithm. To assess best subset, three criteria has to be defined.

i.A search strategy to navigate possible features.
ii.Definition of success and stop criterion of the classifier.
iii.The name of the classifier

The exhaustive search requires high computation power when the number of features is large. Variety of search strategies can be used to navigate the feature subsets. A wide range of search strategies can be used such as greedy search, best-first, branch-and-bound. As a classifier many popular methods can be used such as decision trees, Naive Bayes and support vector machines. "Forward selection" and "backward elimination" are two popular greedy search methods. "Forward selection" start with one feature and adds new features depending on the classification criterion while "backward elimination" starts with all set and progressively eliminates the ones which perform least contribution to the classifier.[22]

Embedded techniques perform feature selection process as part of training process, they do not need divide data into the training and validation sets.

4.1.1.3 Feature Extraction

Methods transform the original dataset with dimension D into a new small dataset with dimensionality *d where d<<D*, while keeping the geometry of original data as much as possible (Figure 4.3).

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} \xrightarrow{\text{feature extraction}} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix} = f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}\right)$$

Figure 4.3: **Feature extraction**: creating a subset of new features by transforming existing features.

In general, neither the geometry of reduced dimension nor the intrinsic dimensionality which means the minimum number of parameters needed to account for the observed properties of the data *d* of the dataset X is known [23]. Therefore, there is no optimum solution on dimensionality reduction process and optimization can be solved by assuming certain properties of the data. The transformation methods may be supervised or unsupervised depending on whether or not output information preserves the classification information.

The original data which will be classified might have redundant information for several reasons:

Firstly, some of the features may have small range and subject to noise so their variation might have redundant information. Secondly, some of the features might have correlation with each other. So the variation information might have the replication of each other. Therefore, in many data preparation cases, it is possible to compact the dimension size without losing informative data by skimming redundant features [24].

In the most learning algorithms, the complexity depends on the number of input dimensions, *D*, as well as on the size of the data sample, N. To improve success of the classification algorithm and reducing the memory usage and computation time, reducing the dimensionality of the problem was sought. Decreasing dimension also reduces the complexity of the inference algorithm during testing [20].

Feature extraction methods can be categorized into two sections namely linear and nonlinear techniques. Linear techniques assume that data lie on or near a linear subspace of the high-dimensional space. Mostly used linear feature extraction techniques are Principal Components Analysis (PCA), factor analysis, and classical scaling [25].

While linear dimension reduction methods assume that distribution of data is linear, the nonlinear techniques have the ability to transform complex nonlinear data. In real world problems data is likely to form a highly nonlinear manifold. Previous studies have proven that nonlinear techniques produce better results on complex artificial data comparing to linear ones. Despite the success of nonlinear techniques on artificial datasets, accuracy of nonlinear dimensionality reduction techniques on real datasets is less convincing.

Maaten branched feature extraction techniques into two main categories namely convex and nonconvex (Figure 4.4). Most of researched techniques are in the convex category in which the objective function is of the form $\varphi(Y) = \frac{Y^T A Y}{Y^T B Y}$ (Rayleigh quotient). The optimal solution for this function can be found by resolving the generalized eigen problem [25].

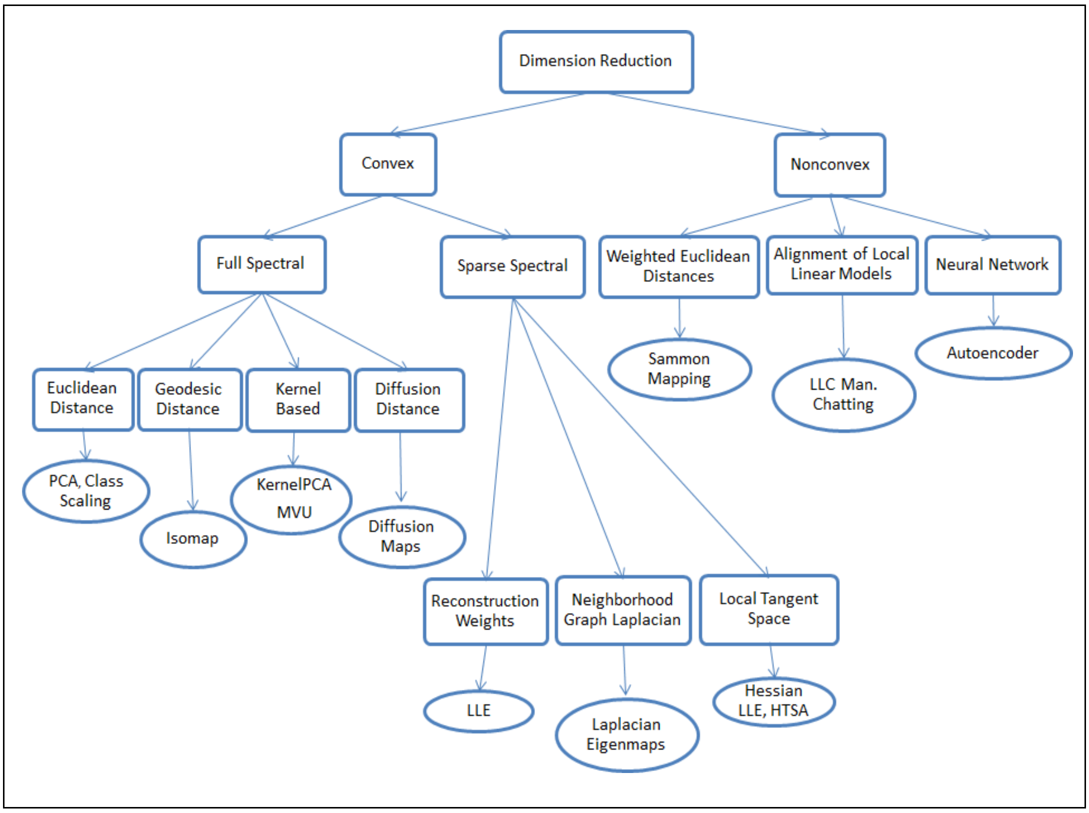


Figure 4.4 Convex techniques resolve the objective functions which do not have any local optima, whereas non-convex techniques resolve objective functions which have local optima

In most cases, the additional information that is lost by discarding some features is compensated by a more accurate mapping in the lower dimensional space. For this reason, feature extraction is commonly limited to linear transforms: $\boldsymbol{y}=\boldsymbol{Wx}$ where y is a linear projection of x. When the mapping is a non-linear function, the reduced space is called a manifold [26].

Osuna categorized the feature extraction techniques in two groups depending on the purpose of the function (Figure 4.5).

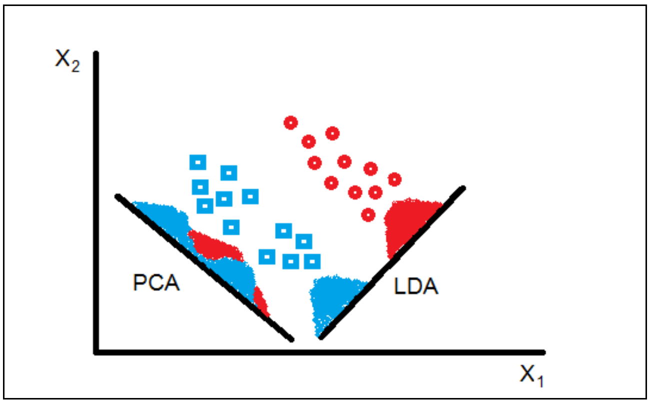


Figure 4.5: Signal representation (PCA) & Classification (LDA)

Signal representation: The objective of the feature extraction mapping is to represent original data in a lower-dimensional space keeping the geometry and the behavior of the data as much as possible (Principal Components Analysis).

Classification: The objective of the feature extraction mapping is to find optimum classification in the lower-dimensional space (Linear Discriminant Analysis).

The main technique of the feature extraction of the linear data is the principal component analysis which performs a linear mapping of the data to a lower dimensional space. PCA works by constructing the correlation matrix of data and computing the eigenvector of the matrix. The eigenvectors which have the largest eigenvalues (the principal components) can be used to reconstruct a large representation of the variance of the original data. The amount of data loss can be adjusted by measuring the weight of the eigenvalues [27].

4.1.2 Data Reduction Techniques Used on Gait Data

The large volume of data resulting from gait analysis is complex, multidimensional and correlated, so the analyzing gait data is a challenging task. The obstacles of analyzing gait data can be summarized as follows.

Firstly, gait data set may have several types of data sources such as kinetic, kinematic, EMG metabolic and anthropometric variables most of which have complex time series data. In terms of data reduction, which gait variables hold the valuable information are not determined in medical domain. Gait analysis systems produce huge amount of information but analyzing these information is limited in clinical experts.

Secondly, as the gait data varies with time, it is difficult to model and the getting the parameters such as peak amplitude, time-to-peak, and mean possibly discard the valuable information of the original data.

Thirdly, In addition to different marker alignments and gait data capturing systems, gait recordings vary intrasubject, intersubject, within trial and between-trial measurements that weakens the statistical conclusions from the gait data. Mostly, the result of different laboratories cannot be compared due to the varying systems, standards and experts.

Lastly, there is no successful metric for gait waveforms to quantatively compare or classify. It is recognized that there is no sufficient and reliable method to reduce gait data and to extract useful information from time variant and correlated gait variables [28].

One of the simplest and most commonly used techniques of analyzing gait data is capturing discrete instants such as peak or base points or certain time events such as toe off , heel rise and toe contact as descriptors of gait pattern [29]. On the other hand, extracting such parameters is subjective and distorts the valuable information of the gait waveforms.

In the study on reviewing the gait data analysis methods, Chau categorized the gait analysis methods into 5 sections namely fuzzy systems, multivariate statistics, fractal dynamics, neural networks and time frequency analysis.

Fuzzy systems partition the continuous time variable into small collection of fuzzy systems. In the classical system membership function (μ) of each small collection has a binary value either 0 or 1 while fuzzy system data point has multiple sets. Although fuzzy systems implemented on continuous signals, temporal distance parameters, a robust fuzzy system can be accomplished by utility of their manipulation through linguistic operators and rules which are not adequately studied in gait analysis. O'Malley et al applied fuzzy K-mean algorithm using two temporal-distance parameters (stride length and cadence) on 88 children with the spastic diplegia which is a form of cerebral palsy and 68 intact children (the control group). Five clusters for the children with cerebral palsy are determined using the information of intact population and cluster validity techniques [30] (Figure 4.6).

V1 is the cluster center of the control group.
V2 is the cluster center of the group which has below normal stride length and compensated an above normal cadence.
V3 is the cluster center of the group which has poor stride length and a below normal cadence.
V4 is the cluster center of the group which has a very low stride length and with a poor cadence.
V5 is the cluster center of the group which has a very low stride length and with a normal cadence.

Malley asserts that having defined this centers, by using four features namely stride length, cadence, leg length and age, it is possible to classify any child with spastic diplegia.

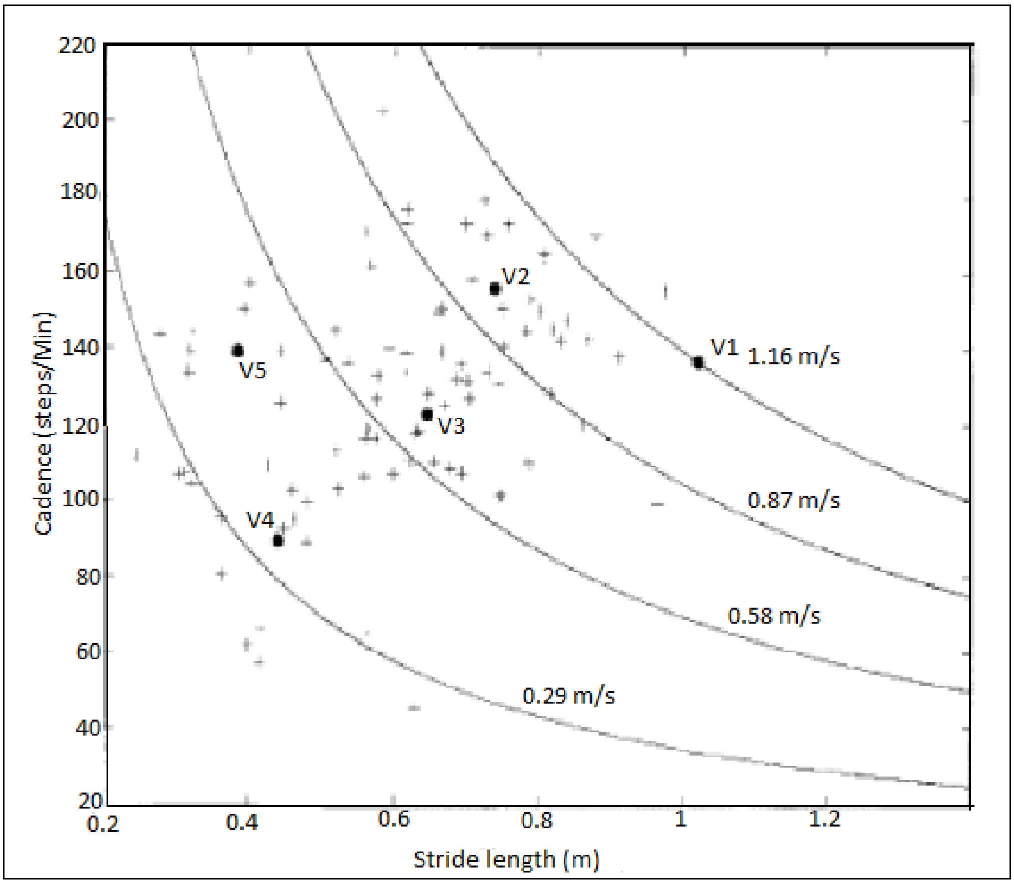


Figure 4.6: Normalized stride length and cadence for children with CP (+), the five normalized cluster centers (•) and constant velocity profiles. (Courtesy to O'Malley [30])

Tan et al used a fuzzy clustering system for diagnosing patients with neurological disorders by processing, extracting and classifying 52 joint angle measurements from still images of 20 human subjects' (about half were patients with Parkinson's disease and the other half were normal) [31]. As a result 84.6% classification accuracy rate was achieved.

Depending on the different usage of the gait signal, PCA has different applications. The first group of study uses condensed version of the gait signal rather than the original signals while the second group uses some gait parameters such as peak powers and energy bursts and the last group uses the entire gait waveform. Classification studies on gait data applying PCA are further explained in the following sections.

Factor Analysis (FA) is another linear multivariate statistical model used in gait analysis. The goal of FA is to find a few orthogonal factors which describe observed variables in terms of a potentially lower number of unobserved variables. FA patterns relationship among variables, reduces data, analyzes latent dimensions, generates factor scores, and test hypotheses. Merkle et al demonstrated that factor analysis can identify coordinated activation patterns across muscles during locomotion. Electromyographic activities of 10 male subjects were taken when they walked on a treadmill. It was revealed that %65 of the variance of seven electromyographic activities received when subjects walked on a treadmill aligned with two orthogonal factors which have offered a unique look at the neuromuscular activity among the musculature of shank, leg, lower back and neck [32].

Correspondence Analysis (CA) is a descriptive and graphical technique for finding correlations among categorical variables. Although CA mainly used in social sciences [33], there have been studies analyzing gait data applying CA. Initially, waveform of gait signals is coded into discrete categories. MCA allows combined analysis of numerous gait variables and can be used to pinpoint relationships among variable for deeper investigation.

Loslever et al proposed a methodology for analyzing six gait signals of fifty subjects with “normal” gait (Figure 4.7). After selecting a set of 25 time windows from the gait data, the range of values within each time window was cut into three fuzzy triangular modalities such as, “low” “medium” and “large”. The resulting table is investigated using multiple correspondence analyses. The MCA model constructed with factor planes of normal subjects is used to compare pathological gait progression during the rehabilitation of a particular subject. [34].

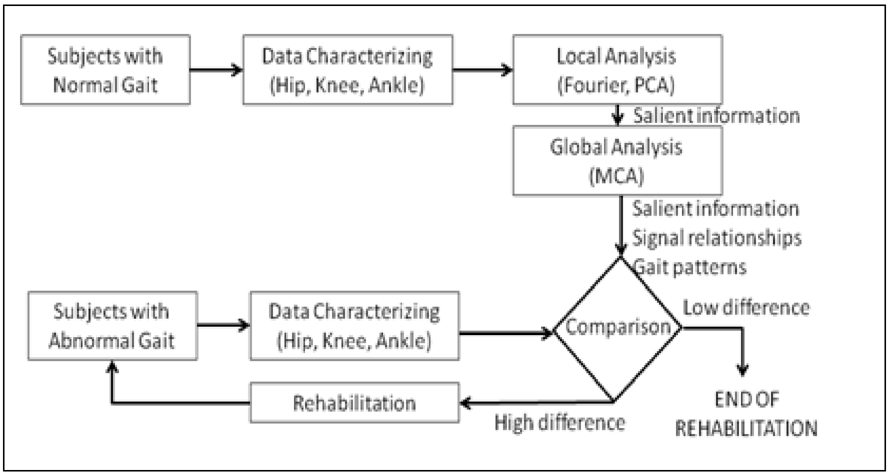


Figure 4.7: Gait analysis methodology

As from statistical point of view, self-similarity implies that statistical properties of the sections express the statistical properties of the body. Objects or data sets which embody self-similarity are broadly termed as ‘fractals’. Fractal dimension exposes the self-similarities and dissimilarities in the signals. Although there are several methods for estimating the fractal dimension of a signal, dispersion analysis and detrended fluctuation analysis have been used on the analysis of univariate gait signals such as stride intervals. However, generally the kinetic and kinematic features have been applied to gait analysis.

Artificial Neural Networks typically consisted of inputs, outputs and hidden layers in between. First the nodes in the hidden layer is trained by adjusting the weights depending on the input and observed outputs. Once the system has been trained, it can predict the outputs by giving the actual inputs. Neural Network (NN) has been used variety of the classification problems including gait analysis. However, processing the raw gait data does not produce accurate results in neural network. Pre-processing and post-processing of input and output variables such as scaling, normalization, Fast Fourier Transforms, rectification and averaging are needed for better performance of NN. Although there have been cases that NN produced better results than classical statistical tools, captured relationships are generally difficult to interpret.

Şen used [13] NN to classify the gait data of 111 patients having knee osteoarthritis with 110 age-matched normal subjects based on the Kellgren-Lawrence scale ("Normal", "Mild", "Moderate", and "Severe"). Three type of feature set has been used in the study:

Feature Set A contained 7 personal features,
Feature Set B contained 8 time-distance features,
Feature set C contained 33 kinetic and kinematic features contained 51samples taken in equally spaced intervals for one gait cycle.

After implementing the feature selection process by using the Mahalanobis Distance criterion, ten different data sets had been created by combining the data points which were obtained by averaging the waveform points and Fast Fourier Transform (Table 4.1). After applying different classification algorithms on these data sets, Backpropagation Neural Network (bpxnc) and Parzen density based (parzendc) classifiers converge faster than the others and best5d data set was selected for further analysis of the gait data.

Table 4.1: List of datasets after feature reduction and selection processes (Şen)

Data Set Name	# of Attributes per Feature	# of Features	Data Set Size	MD
Best51d	51	1	51	14.568
Best25d	25	2	50	18.273
Best10d	10	5	50	23.545
Best5d	5	10	50	19.804
Best2d	2	25	50	20.387
Best1d	1	33	33	14.577
BestFFT25	25	2	50	13.244
BestFFT10	10	5	50	15.589
BestFFT5	5	10	50	18.029
BestFFT1	1	33	33	10.625

In the classification process, five Multi-Layer Perceptrons are trained by five sets composed of different feature vectors and outputs of test set were combined by three different combining rules to reach a final result and accuracy of different combining rules was compared. Additionally, multi-class classification was implemented by combining decision trees which were constructed by using temporal and personal features, and neural networks which were placed on the leaf nodes of the decision tree to predict binary mode classification (Figure 4.8). The final system was tested with test data set and a success rate of around 80% was achieved on the average.

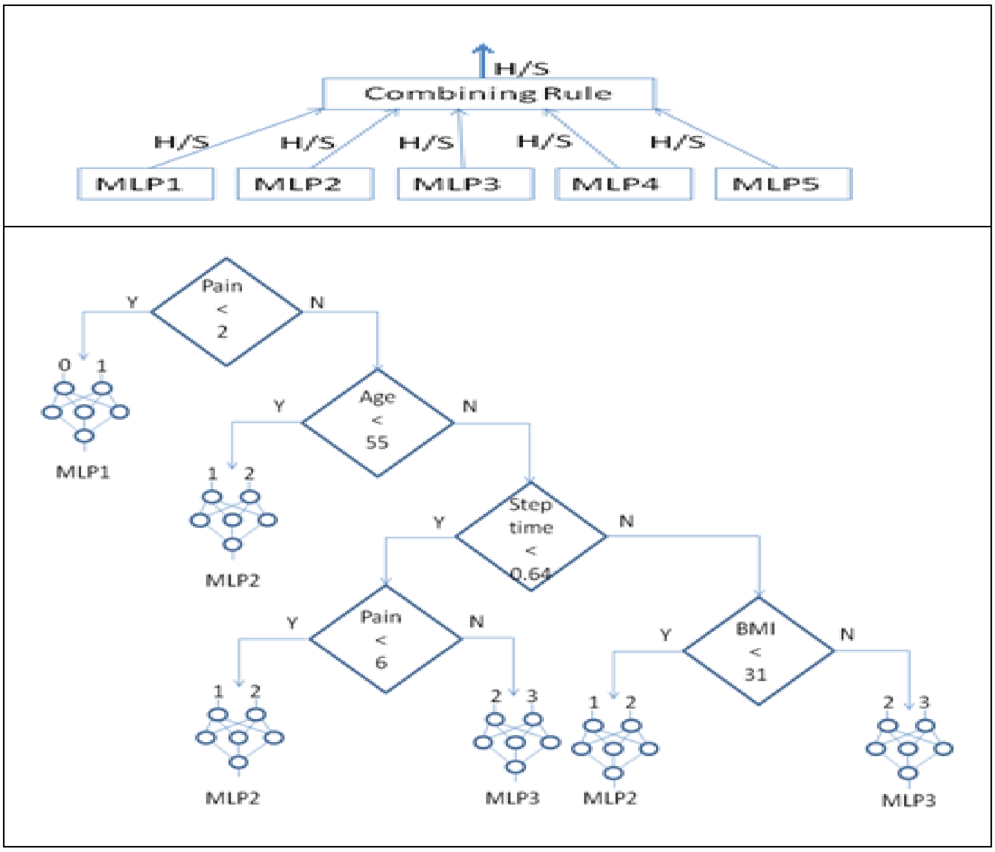


Figure 4.8: (Top) MLP combination schemas. (Bottom) Composed decision tree constructed by personal and time-distance features

Wavelet method is used for signal smoothing and signal discrimination purposes in analysis of gait data. The selection of appropriate wavelet and scaling basis function are the main unresolved questions in the application of wavelet transforms on gait signals [28]. Marghitu et al assert that identifying gait abnormalities by comparing joint rotations of normal and pathological individuals does not allow accurate quantative measure of the disperencies or deviations in the kinematic quantity. They used "wavelets" to decompose the gait waveform of healthy greyhounds and greyhounds with tibial nerve paralysis into its constituent components. By comparing the energy distribution between different components of the transform, the discrepancies between the gait patterns of the normal and paralytic greyhounds were revealed. Significant differences in the signal energy of the tarsal joint angle were discovered. This study exhibit the novel approach for application of wavelet coefficients to demonstrate subtle abnormalities in gait patterns [35].

The discrete wavelet transform decomposes a continuous signal into coefficients which carry spectral and temporal information about the original signal. Tamura classified walking patterns of 20 elderly subjects by using wavelet transform to the acceleration signals obtained by a tri-axial accelerometer attached to the subject while normal walking and up and down walking from a stairway. Tamura used statistical significance difference of two wavelet coefficients ratios to classify walking patterns. The experiments revealed that coefficient ratio of walking up has a statistically significance difference value than the walking down and level walking in anteroposterior direction and that walking coefficient ratio of walking down has statistically significant difference value than the level walking in vertical direction [36].

Table 4.2: Summary of Chau's review on gait data analysis methods [37]
(Data analysis methods arranged by analysis needs)

Processing	Exloratory	Task Driven
Reduction of gait data PCA, FA, Non-linear projection, projection pursuit **Noise removal smoothing** Discrete wavelet transform **Measuring differences in entire gait waveforms** PCA, Fractal Dynamics **Feature extraction** Discrete wavelet transform, PCA	**Finding long-range correlations** Fractal Dynamics **Automatic grouping of subjects** Fuzzy clustering, Self-organizing map **Search for patterns** PCA, FA, MCA, Pattern discovery, decision trees **Relationship between EMG and kinematics** Feed forward neural network, decisionTrees **Studying muscle coordination patterns** Factor analysis	**Modeling relationships among gait variables** Feed forward neural networks, radial basis functions. recurrent neural Networks **Prediction of gait parameters** Feed forward neural network, recurrent neural network **Classification of pathological conditions** Feed forward neural network, decision trees **Confirming a suspected pattern in data** MCA, FA, Decision trees

The summary of Chau's review on the data analysis techniques applied to the gait data is given on Table 4.2. As a conclusion of his review, Chau asserts that no particular method used on gait analysis adequately address all the challenges of quantitative gait analysis.

4.1.3 Studies on Kinetic and Kinematic Features of Gait Using PCA & Kernel PCA

PCA is used by variety of ways on gait data, on the first group of studies, PCA is applied on extracted features of the gait signals rather than directly to whole waveforms. The Second group of studies exploits PCA for the direct dimensionality reduction and interpretation of multiple gait signals. On the last group of studies, PCA is applied to the entire gait waveform [28].

Deluzio used PCA as a data reduction and classification tools to compare 50 patients with end-stage knee osteoarthritis with a group of 63 age-matched asymptomatic control subjects. Gait features used in the study were selected depending on the previous literature relating their relevance to knee OA. Firstly, the group differences of principal component scores of knee flexion angle, flexion moment and adduction moment waveforms were compared. A further analysis was implemented by combining PC's which have statistically significant difference using discriminant analysis [29].

It is known that nature of the gait variables are complex and gait is the result of complex consequence of human locomotion. All the valuable information cannot be described by second-order correlations [37]. Therefore, nonlinear analysis techniques should be used to analyze gait patterns [38]. By mapping nonlinearly separable patterns in the input space to higher dimensional space, they can be classified in a higher dimensional space [39]. Nonlinear mapping to the higher space can be realized by means of a kernel function because of the high computation burden of the mapping to the higher dimension [40]. All the computations of the linear PCA can be done in the input space by applying the kernel trick without mapping to the higher dimension. Consequently, kernel-based PCA (KPCA) may transform the nonlinear pattern of gait data into linearly separable form allowing significant information for gait classification. Wu at al used KPCA to extract the features of 36 discrete parameters (9 spatio-temporal, 27 kinematic parameters) obtained from 24 young and 24 elderly subjects (Table 4.3). The classification performances of the combination of KPCA and Support Vector Machine (SVM) with other gait classifier systems along with PCA-based SVM and original SVM were also compared.

Table 4.3: Combined gait features used in KPCA
HC: Heel Contact HR: Heel Rise TC: Toe Contact TO: Toe Off [38]

Spatio-temporal feature (9)	**Kinematic feature** (27)
Absolute stance, swing and single support duration Normalized stance, swing and single support duration Stride length (m) Gait cadence (step/min) Gait velocity (m/s)	Hip ROM during stance phase, swing phase, HC–TC, TC–HR and HR–TO interval Knee ROM during stance phase, swing phase, HC–TC, TC–HR and HR–TO interval Ankle ROM during stance phase, swing phase, HC–TC, TC–HR and HR–TO interval Hip angle at HC, TC, HR and TO Knee angle at HC, TC, HR and TO Ankle angle at HC, TC, HR and TO

As a result of this study, KPCA and SVM results revealed better performance comparing to PCA-SVM, SVM, ANN and Linear Discriminant Analysis (LDA) on classification of which aimed to distinguish between young and old gait patterns of subjects.

Choosing appropriate kernel function is the key on applying KPCA. Unfortunately, no kernel function can be used for all applications. The selection of the kernel function which generates the best classification pattern depends on the particular task [41].

Deluzio used the entire gait waveform of 27 elderly control subjects and 13 patients to examine whether there is a statistically significant difference between the pre-operative and post-operative gait data [42]. First the PC model has been developed by using the control group's gait data by using flexion angle, adduction moment and flexion moment waveforms. Gait data of the patients were transformed by using the principal component model developed from the normal subjects. After the PC scores (see section 4.4.1) were obtained, Mahalanobis distance of each observation was calculated. Additionally, perpendicular distance of each observation from the hyper-plane defined by the Principal Component Model (PCM) was obtained (Figure 4.9). The PC scores and the residuals compared to %95 confidence limits of the normal subject’s.

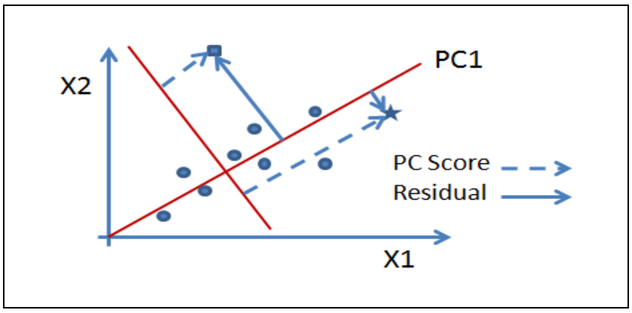


Figure 4.9: The star represents an observation with a large score but small residual whereas the square represents and observation with a large residual but a small PC score

In this study, three patients’ pre and pro-operative gait data were assessed by using the control subjects’ principal component model and checked whether the difference was statistically significant. Mahalanobis Distance criterion was used to measure the distance of each observation from the center of the hyper-plane defined by the PCM for PC scores while sum of square was used to measure the perpendicular distance of each observation from the hyper-plane. Table 4.4 represents outcomes of three patients. According to gait measurements patient P1 improved, P2 got worse and P3 improved. These results were consistent with the clinical assessment measured by Knee Society results as shown in the Table 4.4.

Table 4.4: Results of three patients with respect to Principal Component Model and Knee Society Score.
SP: Sagittal plane FP: Frontal plane TP: Transverse Plane
x indicates significantly different from normal
o indicates that waveform pattern is similar to the normal pattern

Subjects		Knee Society Score / 200	Bone-on-bone forces			Net Reaction Moments			Knee angle	
			SP	FP	TP	SP	FP	TP	SP	FP
P 1	Pre	85	x	o	x	x	x	x	x	x
	Post	185	o	o	o	o	o	o	o	o
P 2	Pre	104	x	o	o	o	x	o	o	x
	Post	99	x	x	o	x	x	x	o	x
P 3	Pre	105	o	x	x	x	x	x	x	x
	Post	138	o	x	x	x	o	x	o	o

4.2 Material and Methods

4.2.1 Review on Dimension Reduction Techniques Used in This Study

4.2.1.1 Principal Component Analysis (PCA)

PCA is an algebraic algorithm that seeks new orthogonal basises or components $\{Z_k\}$ based on the variations of original variables $\{X_i\}$. The principal components Z_k are computed *by* linear combinations of the original variables X_i such that:

$Z_k = a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kp}x_p$ subject to $\sum_i a_{ki}^2=1$

The new basis system in which the original values are transformed are called *principal component (PC)*.PC's are computed in the decreasing order in such a way that the first principal component accounts the most for variation of the original data. As illustrated in Figure 4.10, there is much variation in the direction of P1 which is the first principal component. PCs are artificial variables, so they often have no physical meaning as in original variables. Generally, small number of principal components is enough to explain most of the variation in the original data.

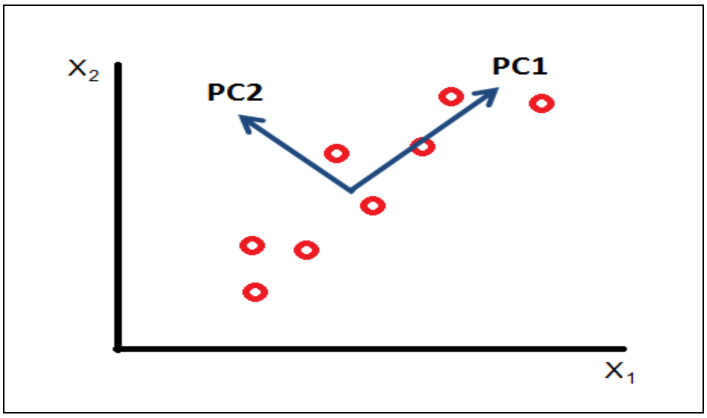


Figure 4.10: PC1 (first principal component) shows the direction of the maximum variance among the data.PC1 and PC2 are orthogonal to each other.

Geometrically, PCA computes best new linear axes which holds the maximum variances of the original data from a space of original dimensionality N to a space of lower dimensionality M, where $M<N$ [28].

Before the application of PCA, the input data is normalized by adjusting data to have zero mean. PCA works by constructing the correlation matrix of data and computing the eigenvector of the matrix. It is mathematically proven that for $k = 1, 2, \cdots, p$ the kth PC score is given by

$Z_k = \varphi_k \mathrm{X}$

where φ_k is the eigenvector of the covariance matrix Σ_x corresponding to its *k*th largest eigenvalue λ_k. Furthermore, if φ_k is chosen to have unit length **($\boldsymbol{\varphi_k^T \varphi_k}$=1)**, then var($z_k$)= λk, where var(z_k) denotes the variance of z_k [43]. Since the eigenvectors are uncorrelated or orthonormal to each other, transformation process can be inverted, X=φZ. This means that original data can be recreated by using eigenvectors and transformed data [29].

By selecting the eigenvectors whose eigenvalues are high (the principal components), new feature space can be rebuilded by preserving much of the variance of the original data. The amount of data loss can be adjusted by measuring the weight of the eigenvalues. Data reduction is accomplished by discarding low explanatory principal components.

Since PC's are computed by the eigenvectors of the covariance matrix Σ_x, orthogonal axes can be found if data presents unimodal Gaussian behavior. If feature data is non-Gaussian or multi-modal Gaussian distribution, PCA will not reveal the variance of data on principal components. PCA can be barely used for classification purposes because it simply rotates all data with the directions of maximum variance. It is not certain that rotated data will hold discriminatory features.

As seen in the left side of the Figure 4.11, x_1 and x_2 highly correlated in the direction of x_2 rather than x_1. After the transformation, it can be clearly seen that z_1 holds much of the variation. The essence of dimension reduction of PCA is that z_2 can be skipped so that the further processes on data can be executed on z_1.

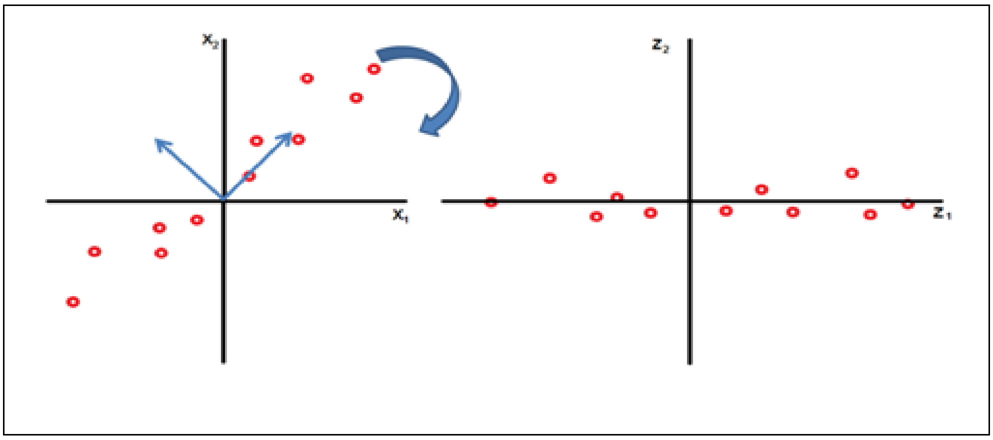


Figure 4.11: Geometrical representation of PCA. Observations on two variables x_1, x_2 and respect to their PCs z_1, z_2.

The mathematical derivation of the Principal Component Analysis method can be seen in Appendix D.

4.2.1.2 Fourier Transform

As sound and light can be represented as wave, many properties of sound are similar to properties of light. Each components of light can be presented by a "pure color" and unique frequency specifications, accordingly a signal can be decomposed into simple periodic sine waves. Analogous to each color of light, each simple component of the signal has specific attribute such as frequency, amplitude and phase (Figure 4.12).

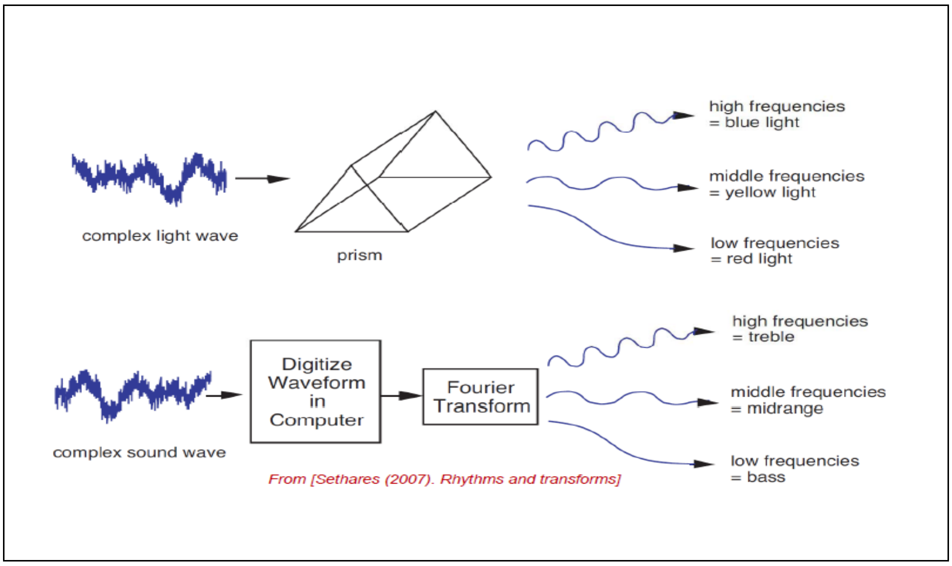


Figure 4.12: Similar to the function of light prism which decomposes the light into different basis colors and frequencies, Fourier Transform decomposes the signal into simple periodic frequencies (Courtesy to Osuna [26])

Mathematically, The Fourier Transform transforms any function into spectrum of different basic sine wave functions each of which has unique characteristics such as specific frequency, amplitude and phase. As analogous to the light prism, Fourier Transform is a tool to obtain the spectrum of simple sinusoids of the given function [44].

A signal is defined as the variation of the magnitude over the time and represented as X(t). On the other hand, a signal can be defined as the function of frequency components and represented as X(f). The collection of values of X(f) at each and every frequency *f* is called the **spectrum** of X(t).The frequency spectrum function is defined as

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt = \langle x(t), e^{j2\pi ft}\rangle$$

The frequency spectrum function X(f) can be converted into time domain function X(t) without any loss. The inverse Fourier transform function which uses the frequency function is defined as

$$X(t) = \int_{-\infty}^{\infty} x(f)e^{j2\pi ft}dt = \langle x(f), e^{-j2\pi ft}\rangle$$

The **Discrete Fourier Transform (DFT)** has the same function as Fourier Transform except that it is applied discrete-time values other than continuous signal.

The mathematical equation is defined as the summation function instead of integral. Additionally, as the frequency samples are discrete, the range of the function can be determined.

Mathematically, the DFT is defined as

$$X(n) = \sum_{k=0}^{N-1} X[k]e^{-j\frac{2\pi}{N}nk} = \langle X[k], e^{j\frac{2\pi}{N}nk} \rangle \ n = 0,1,2, \dots, N-1$$

Similar to Fourier Transform, DFT is computed with the inner product of discrete signal and the sine wave function [45].

4.2.1.3 Downsampling

Downsampling simply refers to reducing the number of samples per second in the time-domain.

It can be observed that downsampling causes the original spectrum to widen on the frequency axis by a factor of m (Figure 4.13). In fact, it exceeds the Nyquist frequency which is the half of the sampling frequency and causes the images of the original spectrum to spill over and fold over to area of interest (aliasing). For a complex signal with a few hundred or thousands of sinusoidal components, downsampling without proper low-pass filtering can cause undesirable distortion. In order to avoid aliasing during downsampling, a low-pass filter is needed to apply to the signal before downsampling or digitization [46]. The low-pass filter allows the frequency components lower than the threshold frequencies so-called cut off frequency in the signal but blocks frequency components higher than the cut off frequency. The frequency components lower than the cut off frequencies are named pass band and the frequency components higher than the cut off frequencies are named stop band. A low pass filter can be very useful as it can eliminate high frequency noise by blocking the high frequency signal and allowing the target frequencies.

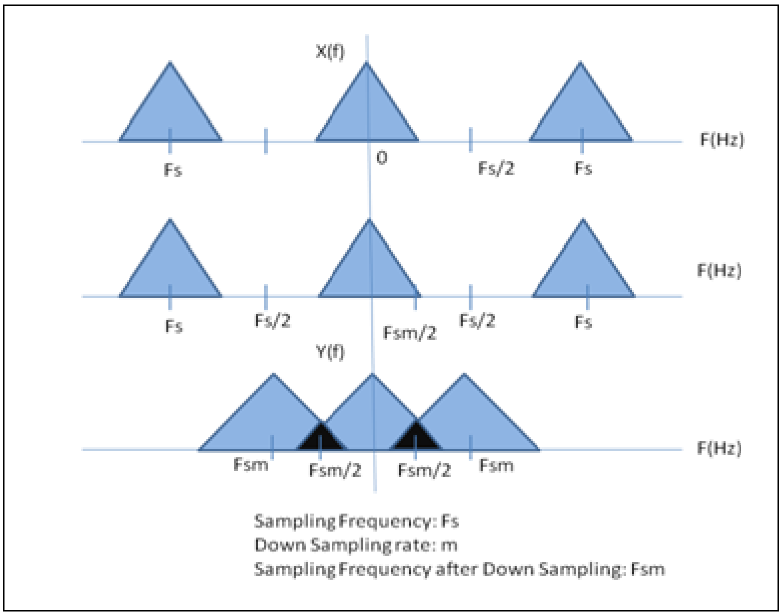


Figure 4.13: Aliasing caused by downsampling

After downsampling the signal with sampling frequency of f_s by a factor of m, the new sampling frequency and Nyquist frequency becomes f_s/m and $f_s/2m$ respectively. The low pass filter must stop the frequency components which are greater than $f_s/2m$ on the original signal to avoid aliasing before down sampling [47].

4.2.1.4 Butterworth Filter

The function of a signal filter depends on the reaction of the frequency response function, H_c. Mathematically, the frequency response function of the Butterworth low-pass filter is defined as

$$|H_c(j\Omega)|^2 = \frac{1}{1 + (\frac{j\Omega}{j\Omega_c})^{2N}}$$

Where j=$\sqrt{-1}$, Ω= the frequency (rad/s), Ω_c= the cutoff frequency (rad/s), and N = the order of the filter. At Ω=0, the frequency response function has no filtering effect. On the other hand, when the frequency band reaches near infinity, the frequency response function becomes 0 and the filter function does not pass any frequency component. Except for frequency value of zero or infinity, filter function partially passes and stops the frequency component (Figure 4.14). When the frequency band is equal to cutoff frequency, the filter function becomes 0.5 for any filter order value.

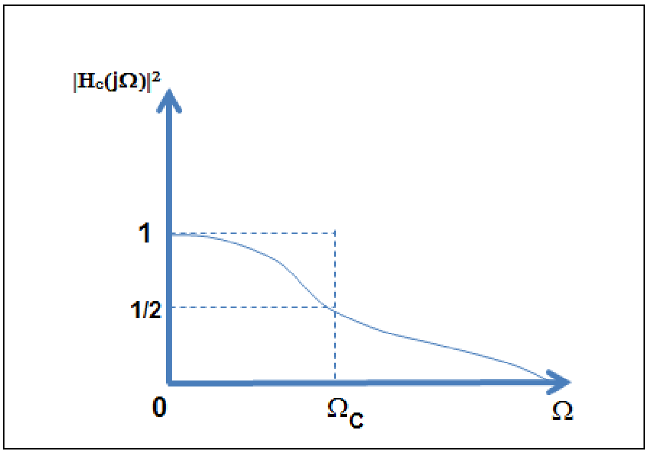


Figure 4.14: Cutoff frequency

Butterworth low-pass filter partially passes the frequency components which are lower or higher than the cutoff frequency. Figure 4.15 shows frequency response. When a higher filter order value is selected, the gradient of reaction of the low pass filter depending on use of variety of the filter orders on the the response function increases. At Ω_c , H becomes 0.707 regardless of the order of the filter [48].

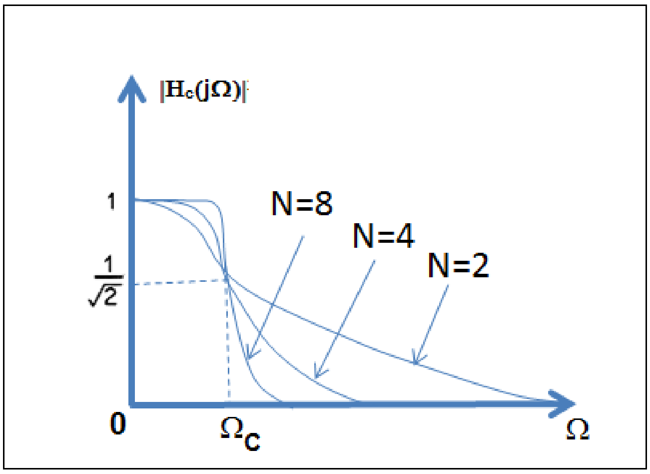


Figure 4.15: Frequency response for the class of Butterworth filters.

4.2.1.5 Sequential Forward Selection (SFS)

Feature subset selection algorithms require an effective search algorithm to pick best subsets navigating through all the possibilities. Selecting the best possible subset requires to explore huge search space. For example, complete search space requires 184,756 feature subsets for the selection of 10 features from 20 features; while complete search space requires 10^{13} feature subsets for the selection of 10 features from 100 feature set. Therefore, an effective strategy is needed to optimally navigate through the search space [49].

Sequential Forward Selection uses basic greedy search algorithm. SFS algorithm begins by selecting one feature and adds a new feature to the feature set when the new feature contributes most to the objective function.

4.2.2 Dimension Reduction Methods Proposed

High dimensional gait data can be reduced using variety of techniques mentioned in previous sections. In this study, two scenarios were implemented to reduce dimension and classify gait data sets. On both scenarios, kinetic, kinematic, time-distance and personal features were used. Each subject has 33 kinetic and kinematic features and each feature has 51 time points which makes total dimension space of 1683. Additionally, 8 time-distance and 4 personal features were used that made 1695 features for each subject.

4.2.2.1 Previous Study

In addition to presented methods, other dimension reduction methods were applied to investigate whether a better classification result could be achieved on this study. In addition to Principal Component Analysis, Linear Discriminant Analysis (LDA) and Kernel Principal Component Analysis (KPCA) were applied as feature extraction methods. Additionally, Mahalonobis Distance and Housedorff Distance methods were applied as feature selection methods to select best features. Because performances of these methods were not satisfactory, these methods were briefly introduced below.

The purpose of the LDA is to find best projection line which has the maximum class seperability. Among the many projection lines which can be drawn, the optimum projection line can be selected by solving the optimization problem of having maximum distance of projected class means (μ_1, μ_2) and minimum projected total variance of the same class observations (Figure 4.16).

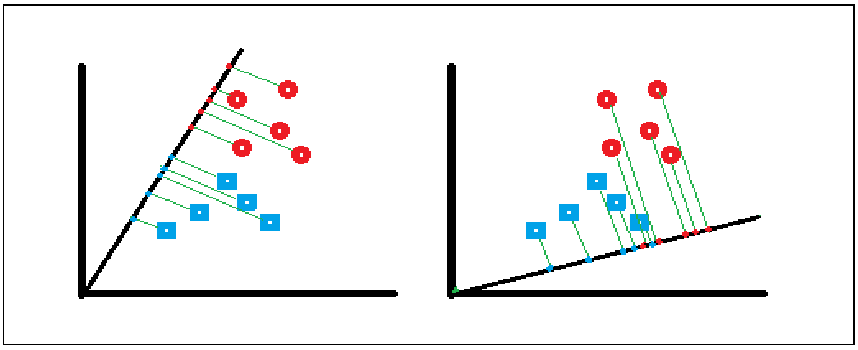


Figure 4.16: Two dimensional two class problem. LDA seeks projection line which holds the optimum class seperability (left)

After using the same data sets and applying SFS, overall classification result of LDA was not satisfactory.

Linear PCA can explain the variance of data which has Gaussian distribution [41]. If data has non Gaussian distribution in input space x_i, we can transform data into higher space $\varphi(x_i)$ and calculate PCA in that higher space (Figure 4.17). The method of calculation of linear PCA in high dimension space is called Kernel PCA. [50].

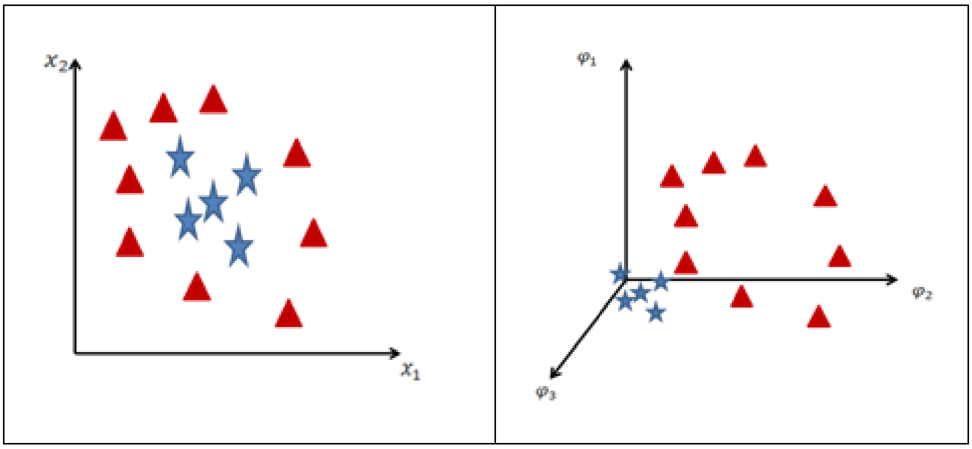


Figure 4.17: The classes on the left are linearly inseparable in the input space. By mapping into higher space, the problem may be linearly separable (Right).

The nonlinear mapping from input space (R^N) to the higher dimensional space (F) is realized by the mapping function (φ) and denoted by $\varphi: \mathrm{R^N} \rightarrow \mathrm{F}, \mathrm{x} \rightarrow \mathrm{X}$ [40].

The KPCA and PCA is the same method in essence, but the difference of KPCA is that calculation of PCA is performed in higher dimensional space *F*. The definition of curse of dimensionality points out that classification power of an algorithm decreases as the number of dimension increases. The idea of mapping data into the higher dimension may seem to be contradictory to the assertion that the performance of the classifier will degrade when the dimension of data increases. However, statistical learning theory states that the contrary can be true. If a simple class decision rule (linear classifier) is used, learning can be simple in high dimensional space [51]. This is a different view of better classification schema that not the curse of dimensionality matters but nonlinearity of data [52]. In that vein, KPCA was used to check whether better classification result could be obtained by moving the gait data into higher dimensional space.

On the other hand, the overall classification success of experiments using KPCA with SVM on different gait data sets was not satisfactory.

In addition to dimension reduction techniques, Mahalonobis Distance and Housedorff Distance methods were used as feature selection methods.

In statistics, Mahalanobis Distance (MD) is a distance metric whose calculation depends on capturing the covariations among the input variables. It differs from Euclidean distance by taking into account the correlations of data. [53]. In some of previous experiments, MD of gait features was calculated and the ones having the greatest MD values were selected for classification by considering that features having the greatest MD might hold maximum class seperability.

Hausedorff Distance (HD) is the "*maximum distance of a set to the nearest point in the other set "* [54]. HD is mainly used to measure the level of similarity of an object which is intermixed within another object. [55].

In previous experiments, six features were selected based on MD and four features were selected based on HD. The selected features were further applied different feature extraction and selection methods.

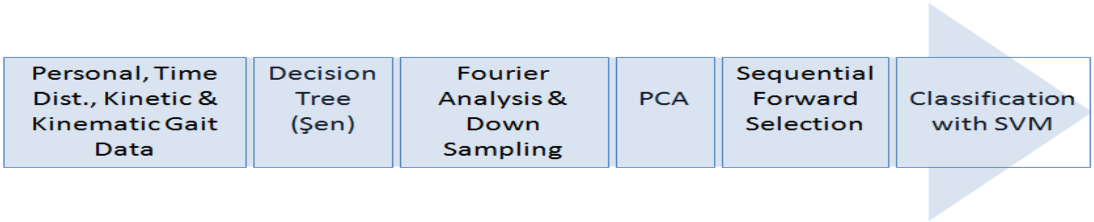
In one of the previous experiments, a new data set was constructed by applying the control group's eigenvector on patient's gait data to increase the difference among the groups of OA. This method is called Best Fit & Point Distribution Model and used by Cootes. The classification result of new data set revealed that classification of normal group is successful. Despite the prior expectations, new data set produced the worst classification accuracy rate (%34.5). The Confusion Matrix of SVM classifier of this data set revealed that 74 subjects out of all 115 subjects were classified as level 0 which means that this method approached all groups into level 0 instead of increasing difference among groups.

4.5.2 Scenario 1

The main objectives of implementing this scenario were to perform combination of signal downsampling techniques and other dimension reduction techniques to reduce data set and to use all gait features not to miss any information that might contribute classification process. Since the kinetic and kinematic features of gait are actually time dependent discrete signals, signal analysis and downsampling methods can be applied in addition to conventional dimension reduction techniques. In this scenario, cadence, walking speed, stride time, step time, single support, stride length, step length, BMI, ache level, first move ache, age and 33 kinetic and kinematic features were used.

Data preparation phase of this scenario was organized by creating four data sets for each level group of OA and by combining 33 features each having 51 data points for each subject. After analyzing waveforms of all subjects, extreme data point values were observed in some cases. Therefore, outlier analysis was performed in such a way that sample data which have more than %30 outlier points of total 51 points were replaced with mean waveform of corresponding level. A point is called an outlier if it has difference of more than 2 standard deviations from the mean. On the first scenario, three different consecutive dimension reduction techniques were applied to gait data set. Downsampling using Fourier analysis with low pass Butterworth filter was the first step of dimension reduction process. As the second step of dimension reduction process, feature based PCA was performed to further reduce the dimension size by selecting PCs having total of 98% eigenvalue threshold meaning that selected PCs explained 98% of the variance of data set. After PCA, six binary classed data sets were constructed by combining each level group (0-1, 0-2, 0-3, 1-2, 1-3, 2-3). As the last part of dimension reduction process, SFS was used on binary classed data sets for sequentially searching the best points (features) which have the most discriminatory effect to the classifier. On the classification phase, six binary data sets each having different features were trained using SVM classifier. As a result, six different classification results were obtained for each subject. The final classification was realized by voting result of six classification output.

4.5.3 Scenario 2



In this scenario, the dimension reduction and classification methods used in Şen's study were combined with the methods used in this study. Şen used decision tree to construct binary classed data sets and MD metric to select best features. In addition to methods used by Şen's, Fourier analysis, PCA and SFS used as dimension reduction methods. In this scenario pain, age, step time, BMI were used on decision tree and KFlex, KMFlex, KPFlex, HPAbd, HFlex and APDor (see Abbreviation Section for detailed names) features were used on binary classification.

As for data preparation phase, a new data set was constructed from the gait database whose data was collected by Şen. Twenty subjects were selected for each level of OA making a data set having total of 80 subjects. In her study, Şen constructed three layered decision tree depending on personal and time-distance features to classify four levels of OA.

As for the next process, same decision tree rules were executed to get binary classed data sets on each of six leaves of decision tree. Since execution of decision tree produced 100% success rate on classifying normal subjects, further implementation were based on the separation among level 1, level 2 and level 3 groups. Therefore, binary classed data sets produced on the leaves were combined into two data sets according to the same binary class groups. Depending on the same feature selection process in Şen's study, two data sets was created with six kinetic and kinematic features which had been selected in that study to classify level 1, level 2 and level 3 groups (1-2,2-3).

The selected features downsampled according to Fourier analysis as implemented in Scenario 1.

In PCA process, PCs were selected with the same 98% eigenvalue threshold. As data sets were binary classed, two separate binary classification sessions were implemented with SVM classifier. Finally, accuracy rates of decision tree and SVM were combined to get the actual success rate.

The differences between Şen's study and this Scenario are that Şen used MD criterion to select the features and 5 data points were selected on each feature. As a result six features each having 5 data points were used on each tree leaf. On the other hand, in addition the feature selection methods used in Şen's study, downsampling with Fourier analysis, PCA and SFS were also used as dimension reduction methods. While Şen used NN on the classification phase, SVM was used in this study.

CHAPTER V

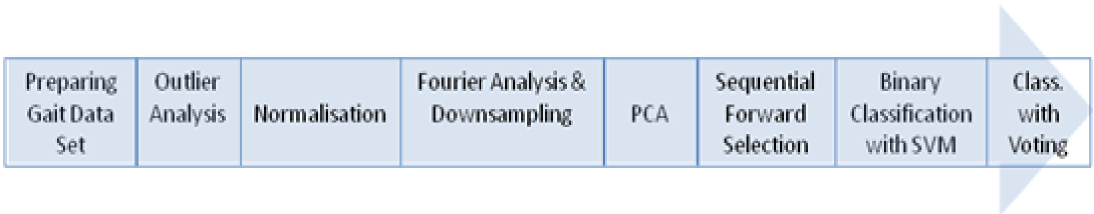
RESULTS

Two different scenarios were implemented to classify the levels of OA by performing different feature extraction and feature selection methods explained in this study. At the beginning of the each scenario, the names of all the methods and processes executed in that scenario were given in horizontally ordered boxes under the title of 'Methods'. Detailed explanation of each process was given in sequentially numbered items. Each scenario actually consists of three main phases which are data preparation, dimension reduction and classification.

Data Preparation, Outlier Analysis, Downsampling According to Fourier analysis processes and PCA methods were realized by Matlab code using Matlab version 9.1 [56] software. SFS and SVM methods were implemented on RapidMiner version 5.1 open source data mining software [57]. Rules of Şen's Decision Tree which was used in Scenario 2 were executed in Microsoft Access using SQL commands. Data sets were hold in MS Excel files and the RapidMiner's Repository tool.

5.1 Scenario 1

Methods



1. Vicon Motion Analysis system produces Microsoft Excel formatted data files for each patient's walking trial. Each subject's data files transferred into four data set files each contained 33 kinetic and kinematic features, 8 time-distance and 4 personal features.

2. During visual analyzing of waveforms, it was detected that some waveforms present exceptional behaviors such as high peaks or sharp slopes. Figure 5.1 shows Hip Moment Abduction waveforms for all the subjects graded as 2. As it can be seen on graphic, waveforms of two subjects can easily separated from the others. To maintain the stability of the calculations, these exceptional cases were eliminated with Outlier Analysis procedure. Firstly, mean and standard deviation of each feature for each level group of OA were calculated. Secondly, any point of all 51 points was considered as an outlier if the difference from this point to the mean was more than 2 standard deviation. If number of total outlier points of a sample were more than 30% of all 51 points, the sample data was replaced with mean of the same level

group of OA. The Matlab Code for the outlier elimination procedure and complete number of subjects found as outlier can be seen in Appendix B.

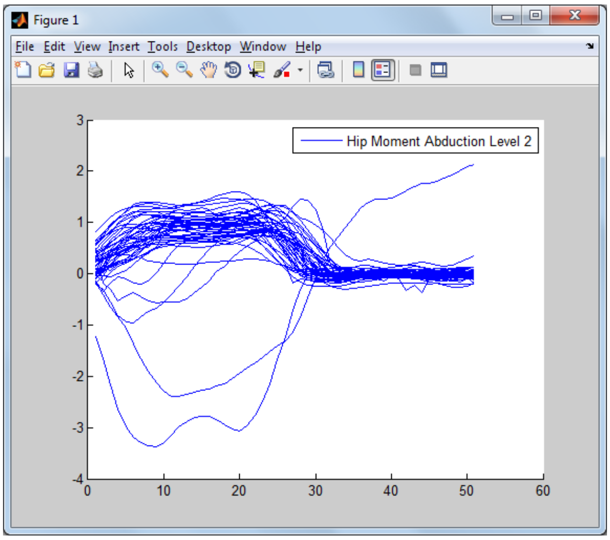


Figure 5.1: Waveforms of Hip Moment Abduction for all subjects graded as 2

3. After the outlier analysis, all samples of four datasets were normalized by subtracting the mean of each feature not to be biased on the high ranged features.

4. The waveformed gait data illustrates the changes of angles of body parts in time according to plane. Therefore, kinetic and kinematic features can be regarded as signals and mathematical signal methods can be applied to these features. After using Fourier analysis to decompose frequency components of a signal, the frequency spectrum can be analyzed to determine at which frequency band much of characteristic of a signal was preserved. When cutoff frequency of a signal is determined, the optimum downsampling rate can be evaluated. As a rule of thumb, the magnitude of –15 db was used as the cutoff frequency band on the frequency axis. As frequency sampling rate was selected as 256 through the experiments, Nyquist frequency rate became half of sampling rate (N/2=128).

   Figure 5.2 shows the frequency spectrum of original Foot Power Rotation (FPro) waveform of sixteen subjects. As magnitude -15 db is matched with frequency band at around 60, it can be concluded that this feature can be downsampled by 2 because this point is roughly half of the Nyquist frequency. This information means that if the original waveform is downsampled by two, the characteristic of the waveform can be generally preserved and new sampling frequency becomes half of the original sampling frequency (f_s). As the Nyquist frequency is half of the sampling frequency ($f_s/2$), Nyquist frequency of the downsampled signal becomes quarter of the original sampling frequency ($f_s/4$). In order to avoid aliasing while downsampling the signal, the frequencies greater than $f_s/4$ should be eliminated. Therefore, a low pass Butterworth filter was used before the downsampling to stop frequencies greater than $f_s/4$ [47]. As it can be seen on the graph on the left, the red line indicates the frequency spectrum after the downsampling of the waveform and blue line presents the frequency spectrum of the original signal.

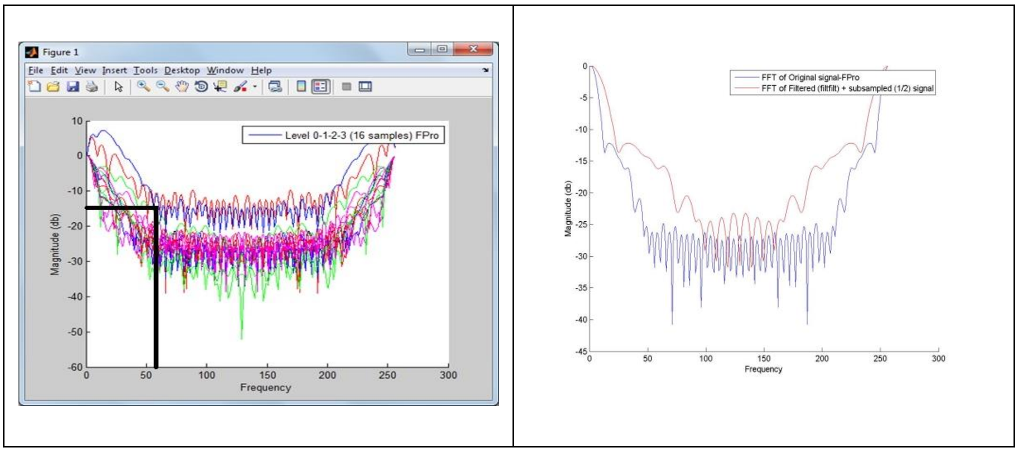


Figure 5.2: Fourier analysis of Foot Rotation

Another example of Fourier analysis is shown in Figure 5.3. As seen on the graph, magnitude of -15 is matched approximately at frequency band of 55 which is around half of Nyquist rate (128). That means that if the original waveform is downsampled by 2, the characteristic of the waveform can be generally preserved. To downsample the signal by 2 without aliasing, frequencies components greater than the 1/4 of sampling frequency should be stopped by using low pass Butterworth filter.

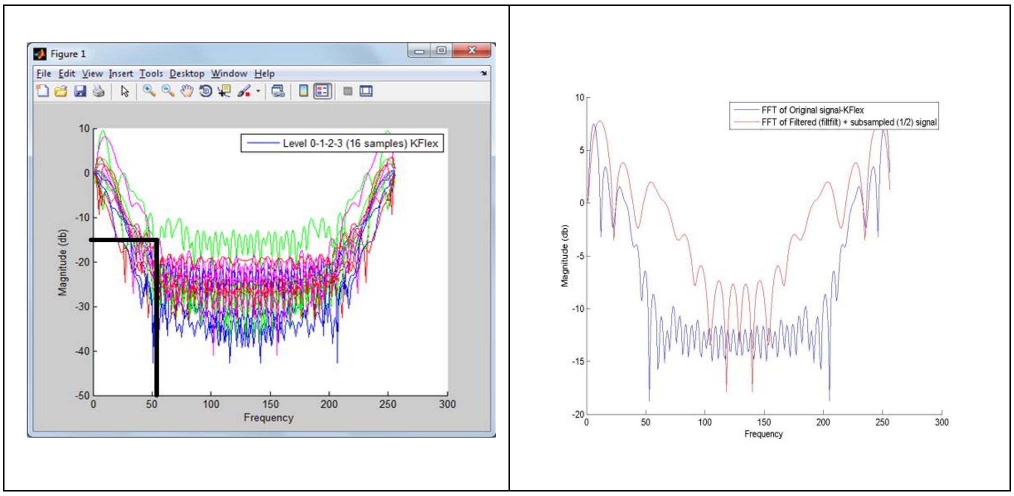


Figure 5.3: Fourier analysis of Knee Flexion

After Fourier analysis was performed for 33 features and their downsampling rate and low pass Butterworth filter parameters were determined, all features were downsampled according to these specifications. The complete list of downsampling rates of all features are listed in Table 5.1 The Fourier frequency spectrum analysis for each feature and Matlab implementation code can be accessed in Appendix C.

Table 5.1: Downsampling rates for each feature after applying Fourier analysis

Feature	Downsampling Rate	Feature	Downsampling Rate	Feature	Downsampling Rate
FPro	1/2	KPVal	1/1	KMFlex	1/2
HAbd	1/2	KPRot	1/1	FMAbd	1/1
FRot	1/2	KRot	1/2	FMRot	1/2
KFlex	1/2	HFlex	1/2	HPTot	1/1
FDor	1/2	PObliq	1/2	HPFlex	1/1
HRot	1/2	PTilt	1/2	HPAbd	1/1
KMRot	1/1	PRot	1/1	HPRot	1/1
KMVal	1/1	KVal	1/2	APTot	1/1
FMDor	1/1	HMRot	1/2	APDor	1/1
KPTot	1/1	HMAbd	1/1	APAbd	1/1
KPFlex	1/1	HMFlex	1/1	APRot	1/1

5. After downsampling each feature according to Fourier analysis, 33 feature data sets with different dimension sizes were obtained. For example, Foot Rotation Power and Knee Flexion feature data set were reduced to size of 26 (from 51, hence ½ downsampling) according to their downsampling rates. Before applying PCA, four levels of each feature data sets were combined into one data set and each feature data set was applied to PCA and PCs having total of %98 eigenvalue threshold were selected. By computing the inner product of selected PCs and input feature set, new points were calculated. The number of PCs selected after %98 threshold for each feature are listed in Table 5.2.

Table 5.2 Feature and selected number of PCs with %98 threshold

Feature	Number of PCs Exceeding %98 Threshold
FRot	2
FPro	4
HFlex	4
HRot	4
KRot	5
KVal	5
HAbd	5
KFlex	5
HMAbd	6
HMFlex	9
FDor	7
KMRot	6

KMVal	4
FMDor	4
KPTot	16
KPFlex	16
KPVal	15
KPRot	15
PObliq	4
PTilt	2
HMRot	6
KMFlex	6
FMRot	6
PRot	4
FMAbd	4
HPTot	15
HPFlex	15
HPAbd	10
HPRot	13
APTot	11
APDor	10
APAbd	7
APRot	14
TimeDist & Pers.	3
Total	262

After PCA was applied and number of PCs was determined for each feature, selected files were combined and six binary data sets were created according to levels (0-1, 0-2, 0-3, 1-2, 1-3, 2-3).

On the last part of dimension reduction phase, feature selection method was applied to further reduce 262 attributes for each of six data sets. Forward Selection algorithm was used as for feature selection method. Forward Selection works by adding a new attribute on each step according to search algorithm and by examining temporary data set according to specified classifier. If any of the features increased the classification performance, that feature was added to the selection list. SVM classifier was used as the classification criterion of feature selection method. There are various feature selection methods available in RapidMiner software. Benjamin Schowe discussed feature selection methods available in RapidMiner software in a study called “Feature Selection for high-dimensional data with RapidMiner”. To use the best feature selection methods, various feature selection algorithms were tested in binary data set of “2-3” among the algorithms used in Schowe’s study. Table 5.3 shows the algorithms and their success rates for the same data set. As the best result was achieved on “Forward Selection” method of tested feature selection algorithms, the forward selection method was used on the rest of the studies.

Table 5.3 Performance of different feature selection methods

Feature Selection Method	Number of Selected Features	Classification Success Rate of Level Group 2-3
RFE	**3**	**%54**
RCCW	**11**	**%67**
MRMR-FS	**107**	**%52**
Forward Selection	**5**	**%76**
Backward Elimination	**242**	**%50**

RFE: Recursive Feature Elimination
RCCW: Recursive Conditional Correlation Weighting
MRMR-FS: Minimum Redundancy Maximum Relevalence Feature Selection

6. On the last step of Scenario 1, the reduced binary classed data sets were classified using SVM classifier tool (LibSVM) of RapidMiner. The parameters of the SVM classifier are given in Table 5.4.

Table 5.4 Configuration of Rapidminer SVM classifier

Software tool	RapidMiner LibSVM
SVM type	nu-SVC
Kernel type	rbf
Gamma	0.0
Epsilon	0.001
nu	0.4

Rapid Miner’s validation tool was used on training and testing of all data sets by choosing 10 fold cross validation. Ten fold cross validation test method splits data set into ten subsets reserving one subset for testing on each round of ten executions.

Table 5.5 shows confusion matrix of six binary classed data sets. The numbers under the columns titled as **True** indicate the number of subjects that actually belonged to the level group of that column and numbers across the row titled as **Pred.** indicate number of subjects predicted with the classifier. The number at intersection of the same numbered column and row titles indicates the number of accurate classified subjects. For example, the number at the intersection of column **True 1** and row **Pred 1** indicates the accurate number of subjects that the classier predicted as 1 for level group 1. The row titled as **Class Recall** indicates the accuracy of classifier for each level group. For example, class recall of 95% under the column titled **True 1** presents the accuracy of the classifier for level 1. The column titled as **Class Precision** indicates the accuracy of classifier for a specific level group that the classifier predicted. For example, class precision of 97.87% on row titled as **Pred. 0** indicates the accuracy of the classifier for all the subjects that the classifier labeled as level 0. The row called **Selected Features** presents the name of the PCs used on the classification process. What is the effect of selected PCs is discussed on Appendix E .

Table 5.5 Confusion matrix of binary data sets

Level 0-1 Confusion Matrix			
	True 0	**True 1**	**Class Precision**
Pred. 0	46	4	92.00%
Pred. 1	0	16	100.00%
Class Recall	100.00%	80.00%	
Selected Features	PObliq PC2, Time-Dist. & Personal PC1, HAbd PC4, KPFlex PC6		

Level 0-2 Confusion Matrix			
	True 0	**True 2**	**Class Precision**
Pred. 0	46	1	97.87%
Pred. 2	0	68	100.00%
Class Recall	100.00%	98.55%	
Selected Features	Time-Dist. & Personal PC1,Time-Dist. & Personal PC2, KFlex PC5, HMAbd PC2		

Level 0-3 Confusion Matrix			
	True 0	**True 3**	**Class Precision**
Pred. 0	46	1	97.87%
Pred. 3	0	44	100.00%
Class Recall	100.00%	97.78%	
Selected Features	Time-Dist. & Personal PC1, Time-Dist. & Personal PC2, FMDor PC1		

Level 1-2 Confusion Matrix			
	True 1	**True 2**	**Class Precision**
Pred. 1	11	1	91.67%
Pred. 2	9	68	88.31%
Class Recall	55.00%	98.55%	
Selected Features	HPRot PC6, FMAbd PC2, PTilt PC1, PRot PC1		

Level 1-3 Confusion Matrix			
	True 1	**True 3**	**Class Precision**
Pred. 1	16	2	88.89%
Pred. 3	4	43	91.49%
Class Recall	80.00%	95.56%	
Selected Features	KMFlex PC4, FMAbd PC2		

Level 2-3 Confusion Matrix			
	True 2	**True 3**	**Class Precision**
Pred. 2	62	17	78.48%
Pred. 3	7	28	80.00%
Class Recall	89.86%	62.22%	
Selected Features	KRot PC4, Time-Dist & Personal PC1,POblique PC2, PTilt PC2		

Once SVM classifiers trained for six binary classed data sets, each level of data set was classified using each of six trained data sets. As shown in Table 5.6 each sample was tested on six binary data sets. As a result, six results were obtained for each sample and majority voting method was used to get the final result. The one which has more than two same classes output wins the majority vote. If there is a tie among classification results, the level having the maximum sum of confidence (probability estimates which is a function of distance to the separating hyperplane) wins the voting. The overall classification result for 80 subjects was presented in Table 5.7.

Table 5.6 Majority Voting results of each level. The number on each cell indicates the classification result and the number given in parentheses indicates the confidence

Majority Voting Results For Level 0						
0-1	**0-2**	**0-3**	**1-2**	**1-3**	**2-3**	**Voting Result**
0	0	0	2	1	2	0
0	0	0	2	1	3	0
0	0	0	2	1	2	0
0	0	0	2	1	2	0
0	0	0	2	1	2	0
0	0	0	2	1	2	0
0	0	0	2	1	2	0
0	0	0	2	3	2	0
0	0	0	2	3	3	0
0	0	0	2	1	3	0
0	0	0	2	3	3	0
0	0	0	2	3	3	0
0	0	0	2	3	2	0
0	0	0	2	3	2	0
0	0	0	2	3	2	0
0	0	0	1	3	2	0
0	0	0	2	3	2	0
0	0	0	2	3	2	0
0	0	0	2	3	2	0
0	0	0	1	1	3	0

Majority Voting Results For Level 1												
0-1 Output/ Confidence		**0-2** Output/ Confidence		**0-3** Output/ Confidence		**1-2** Output/ Confidence		**1-3** Output/ Confidence		**2-3** Output/ Confidence		**Voting Result**
1	0.73	2	0.62	3	0.53	2	0.73	1	0.71	3	0.60	1
1	0.51	2	0.51	0	0.51	2	0.72	1	0.52	3	0.59	2
1	0.63	2	0.77	3	0.71	2	0.68	1	0.71	2	0.65	2
1	0.68	2	0.81	3	0.71	2	0.72	1	0.73	2	0.73	2
1	0.54	2	0.61	3	0.56	2	0.73	1	0.60	3	0.72	2
0	0.75	2	0.54	3	0.51	1	0.73	1	0.73	2	0.69	1
0	0.75	2	0.51	0	0.51	1	0.72	1	0.79	2	0.61	1
1	0.62	2	0.57	0	0.53	2	0.73	1	0.74	2	0.84	2
1	0.71	2	0.78	3	0.76	1	0.71	1	0.73	2	0.81	1
1	0.55	2	0.77	3	0.76	1	0.73	1	0.73	2	0.83	1
1	0.73	2	0.89	3	0.80	2	0.70	3	0.68	2	0.73	2
1	0.73	0	0.51	0	0.53	2	0.70	1	0.73	2	0.72	1
1	0.73	2	0.58	3	0.53	1	0.73	1	0.64	2	0.70	1
1	0.73	0	0.74	0	0.77	1	0.69	1	0.73	2	0.69	1
1	0.68	2	0.62	0	0.78	1	0.73	1	0.73	2	0.63	1
1	0.73	2	0.85	3	0.77	1	0.73	1	0.73	2	0.65	1
1	0.78	0	0.54	0	0.64	2	0.73	1	0.66	2	0.65	1
1	0.73	2	0.51	0	0.59	2	0.73	1	0.72	2	0.57	2
1	0.73	2	0.82	3	0.81	2	0.73	1	0.62	3	0.55	2
1	0.73	0	0.51	0	0.51	2	0.72	1	0.55	3	0.65	1

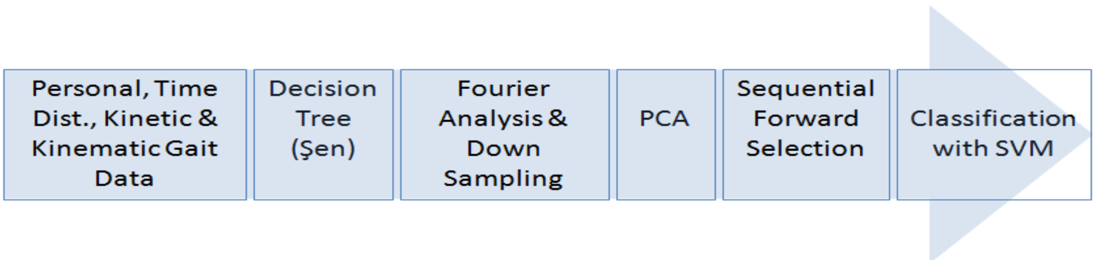
Majority Voting Results For Level 2						
0-1	**0-2**	**0-3**	**1-2**	**1-3**	**2-3**	**Voting Result**
1	2	3	2	1	2	2
1	2	3	2	3	2	2
0	2	3	2	1	2	2
1	2	3	2	3	2	2
1	2	3	2	3	2	2
1	2	3	2	1	2	2
1	2	3	2	3	2	2
1	2	3	2	3	2	2
0	0	0	2	3	2	0
1	2	3	2	3	2	2
0	2	3	2	3	2	2
1	2	3	2	1	2	2
1	2	3	2	1	2	2
1	2	3	2	3	2	2
0	2	3	2	1	2	2
1	2	3	2	1	2	2
1	2	3	2	3	2	2
0	2	3	2	1	2	2
0	2	3	2	1	2	2
0	2	3	2	1	2	2

Majority Voting Results For Level 3 (Number in parentheses indicates confidence value)						
0-1	**0-2**	**0-3**	**1-2**	**1-3**	**2-3**	**Voting Result**
1	2	3	2	3	3	3
0	2(0.55)	3(0.54)	2(0.73)	1	3(0.66)	2
0	2	3	2	3	3	3
1	2	3	2	3	2	2
0	2	3	2	3	3	3
0	2	3	2	3	3	3
0	2	3	2	3	3	3
1	2	3	2	3	3	3
1	2	3	2	3	3	3
1	2	3	2	3	3	3
1	2	3	2	3	2	2
0	2	3	2	3	3	3
0	2	3	2	3	3	3
0	2	3	2	3	3	3
0	2	3	2	3	3	3
0	2	3	2	3	3	3
0	2	3	2	3	3	3
0	2	3	2	3	3	3
1	2	3	2	1	2	2
0	2	3	2	3	3	3

Table 5.7 Overall classification result

Voting Results				
	Level 0	**Level 1**	**Level 2**	**Level 3**
Success Rate	%100	%60	%95	%80
Overall Success Rate	%83.75			

5.2 Scenario 2
Methods



In this scenario, the decision tree implemented in Şen's study was used with the combination of dimension reduction techniques and performances of dimension reduction techniques were evaluated.

1. A new data set was constructed by selecting 20 subjects for each level of OA making a data set of total 80 subjects as in Şen's study.

2. As the next step of Scenario 2, decision tree rules which were already implemented in Şen's study were executed. This step can be regarded as both data preparation and classification phase. In her study, Şen constructed three layered decision tree depending on personal and time-distance features to classify four levels of OA. On each one of six leaves of the decision tree, a binary classed data sets were obtained and binary input multi layered perceptron was used to finalize the classification (Figure 5.4). In this study, the same decision tree rules were applied to 80 subjects' gait data and six binary classed data sets were obtained on the leaves of the decision tree. As the decision tree succeeded to classify all normal subjects and binary classed data sets were obtained on the leaves, one data set holding level 1 and level 2 groups with the number of 22 subjects and one data set holding the level 2 and level groups 3 with the number of 31 subjects were created. Table 5.8 lists the number of correct and incorrect predictions for each leaf. The underlined numbers indicate incorrect predictions.

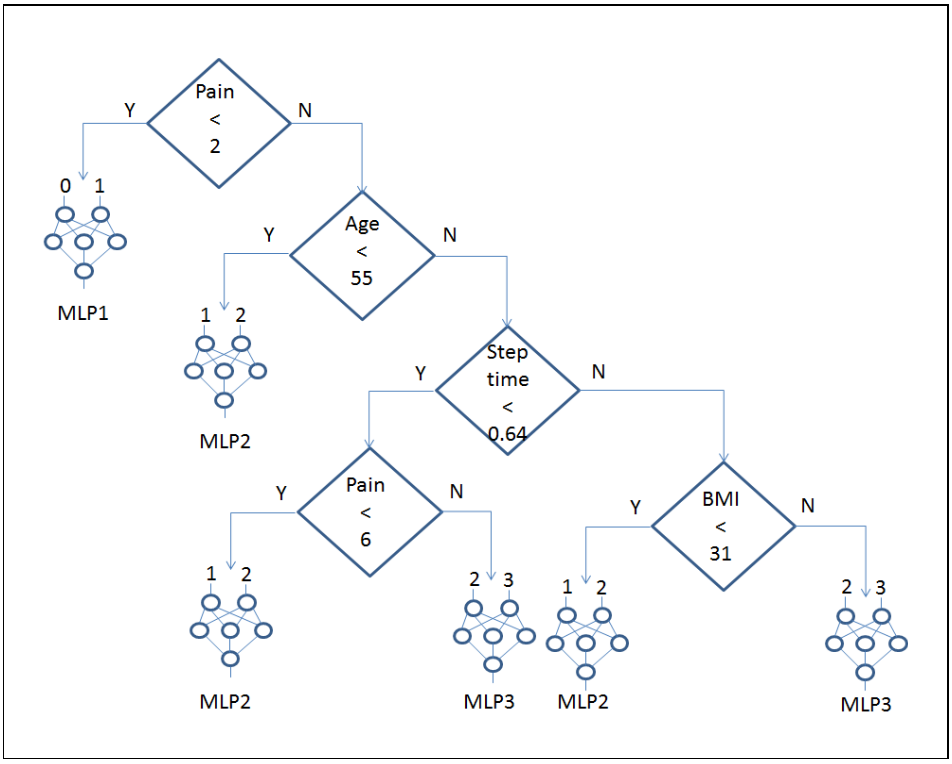


Figure 5.4: Decision Tree Implemented by Şen

Table 5.8 Performance of decision tree

Number of Subjects On Each Level of OA For Each Leaf After the Decision Tree Prediction (Underlined Numbers Indicate Incorrect Predictions)							
Level of OA	**Leaf 1 (0-1)**	**Leaf 2 (1-2)**	**Leaf 3 (1-2)**	**Leaf 4 (2-3)**	**Leaf 5 (1-2)**	**Leaf 6 (2-3)**	**Total**
0	20	0	0	0	0	0	20
1	0	9	3	<u>4</u>	3	<u>1</u>	20
2	0	5	2	9	0	4	20
3	0	0	0	6	<u>2</u>	12	20

3. As for the data selection process, the same features used as in Şen's study. Şen used MD criterion to select the best data set from the data sets created by combination of features. Six features which were used in Şen's study were preselected in this scenario. KFlex, KMFlex and KPFlex features were selected for the separation of level 1 and level 2 and HPAbd, HFlex and APDor features were selected for the separation of level 2 and level 3 as in Şen's study. By combining the same class binary data sets, 1-2 group data sets hold 22 patients and 2-3 group data set hold 31 patients.

4. On the next step of dimension reduction phase, selected features were downsampled according to Fourier analysis rates as computed in Scenario 1. Accordingly, KMFlex, KFlex and HFlex were downsampled by ½ and no downsampling applied on KPFlex, APDor and HPAbd.

5. On next dimension reduction phase, PCA was applied. Number of PC's with selecting total %98 eigenvalue threshold is listed in Table 5.9

Table 5.9 Number of PCs

Feature	**Number of PCs with %98 Threshold**
KFlex	5
KMFlex	5
KPFlex	9
HFlex	4
APAbd	6
APDor	9

After PCA was applied, feature set of level 1-2 and 2-3 classifiers were reduced to size of 19.

6. On the last dimension reduction phase, two data sets were applied to feature selection process using binary SVM classifier. Finally, dimension of feature set classifier 1-2 was reduced to 3 including the class attribute and dimension of feature set having class 2 and 3 was reduced to 3 including the class attribute.

7. On the classification phase, data sets were classified using SVM classifier tool of RapidMiner. Finally, decision tree results and SVM results were combined. The combined accuracy results are presented in Table 5.11. Underlined numbers indicate incorrect predictions of decision tree and addition of numbers indicates the numbers gathered from the different leaves of decision tree and SVM classifier. Row titled as **Class recall** shows the overall performance for each level of OA.

Table 5.10 Final Confusion Matrix

Final Confusion Matrix of Decision Tree + SVM (Underlined Numbers Indicate The Incorrect Results of Decision Tree)					
	True 0	**True 1**	**True 2**	**True 3**	**Class precision**
Predicted 0	20	0	0	0	100%
Predicted 1	0	11	4+<u>5</u>	0	55%
Predicted 2	0	2	12+5	1	80%
Predicted 3	0	0	5+<u>2</u>	13	60%
Class Recall	100%	55%	85%	65%	

The final comparison of accuracy rates of Scenario 2 with Şen’s study is given in Table 5.12. Together with Scenario 1 and Scenario 2 the difference between Şen’s result and this study is mainly on level 0 and level 1.

Table 5.11 Comparison of Performances with Şen’s study

	Scenario 1	Scenario 2	Şen’s Study
Level 0	100%	100%	90%
Level 1	60%	55%	75%
Level 2	95%	85%	85%
Level 3	80%	65%	70%
Overall	83.75%	76.25%	80%

CHAPTER VI

CONCLUSION

6.1 **Results**

Since one of the main problems of gait analysis is having high dimensional and correlated data with low number of patient data, dimension reduction process is the key for successful classification by optimally preserving the valuable data. Therefore, different scenarios were tested in this study by using combination of different feature selection and feature extraction methods and by generating different data sets to improve the accuracy of classifiers.

After analyzing the graphs of the gait waveforms on each level group, it was observed that some of the gait trials have extreme values which might bias the calculations. To eliminate the distraction of extreme results, outlier analysis was performed to gait data sets before the data reduction process.

On the implementation of dimension reduction phase, downsampling with Fourier analysis techniques, PCA and SFS feature selection methods were implemented.

After dimension reduction phase, classification process was performed as the last step of scenarios. SVM classifier was used in all scenarios, because it is regarded as a well established classification method in pattern recognition community. Majority voting and decision tree were also used on classification phase.

On the first scenario, all kinetic and kinematic gait features and time-distance features were used not to miss any valuable information. As the waveformed gait data can be accepted as a discrete signal, signal processing techniques were applied to reduce the time samples of the signal while keeping the maximum information. Downsampling of the signal was realized by applying a low pass filter according to Fourier analysis result on the first step of dimension reduction phase. In addition to signal reduction technique, PCA is executed by keeping the PCs having total 98% eigenvalue threshold. As the last step of dimension reduction, forward selection method of Rapid Miner was used to select best attributes which contribute most to the accuracy of the classifier. Gait data was split into six binary data sets and dimension reduction processes were applied step by step. When data reduction processes were applied, SVM classifier were trained using the six binary data sets and data of 80 subject were examined on each of binary data sets. After the voting process, the final classification result was obtained. As a result, % 83.75 success rate which was slightly better than Şen’s study was achieved.

On the second scenario, the same data preparation and feature selection path was followed as Şen implemented in her study. On the other hand, some of the dimension reduction techniques were also applied such as downsampling according to Fourier analysis, PCA and SFS.As a result %76.25 overall success rate was obtained.

Despite different dimension reduction and classification methods were implemented in scenario 1 and scenario 2, approximately same results were obtained (scenario 1: %83.75, scenario 2: 76.25). In both scenarios control group were classified with % 100 accuracy rate. The classification success rate order from maximum to minimum in both methods was in the sequence of level 0, level 2, level 3 and level 1.

The worst accuracy rate was obtained on level 1 (scenario 1: %60, scenario 2: %55). There may be two reasons behind the low accuracy rate of level 1. Firstly, only twenty subjects were used on each level group and level 1 group is minimum of all level groups that might not enough to represent characteristic of level 1. Secondly, Kellgren & Lawrence grading system is assumed as the main tool in classification of osteoarthritis (OA) in medical domain. On the other hand, there are spectism on OA community over the lack of standardization on Grade 1 OA. Some of studies on this matter suggest that Grade 1 groups either should be treated as a sub group or excluded from the analysis of knee OA. [59] The classification result of this study supports inconsistencies of Grade 1.

Additionally, many studies have criticized Kellgren & Lawrence grading system because of limitation of radiographic OA. Despite, there are more advanced examination techniques other than X-Ray, they are not ubiquitous. [60]

After result of all dimension reduction phases, Time-distance & Personal Features, POblique, HAbd, KPFlex, KFlex, HMAbd, HMDor, HPRot, FMAbd, Ptilt, Prot, KMFlex, FMAbd, Krot were selected. Almost all of the features selected in this study were also used in Şen's study and these features were among the ones which had highest MD values. The comparison of the features used on both studies was given in Table 6.1.

In group 0-1 all lower body parts affected so knee, hip and pelvis related features appeared in both study.

In group 0-2 Şen used almost the same body part as in group 0-1. In this study knee and hip related features were used.

In group 1-2 Şen mostly used knee related features. On the other hand, hip and pelvis related features were used in this study.

In group 1-3 Şen mostly used knee related features while in this study a knee related and a foot related features were selected.

In group 2-3 Şen mostly used hip related features while knee related and pelvis related features were used in this study.

While Şen used direct feature data, PC scores of features were used in this study.
Şen used 0-1,1-2 and 2-3 binary data sets. On the other hand 0-1, 0-2, 0-3, 1-2, 1-3 and 2-3 binary data sets were used in this study.

Table 6.1 Feature selection

	Features Used By Şen (Time-Distance & Personal features used in decision tree)	**Features Used in Scenario 1 (PCs)**
0-1	FMDor, HFlex, KMFlex, KFlex, PTilt, FRot	PObliq, HAbd, KPFlex, Time-Dist. & Personal
0-2	FMDor, KFlex, HFlex, FDor, KRot, PTilt	KFlex, HMAbd, Time-Dist. & Personal
1-2	KFlex, KMFlex, KPFlex, HPTot, KPTot, KVal	HPRot, FMAbd, PTilt, PRot
1-3	KPFlex, KPTot, KFlex, PObliq, KMRot, KVal	KMFlex, FMAbd
2-3	HPAbd, HFlex, APDor, HRot, HPFlex, KVal	KRot, PObliq, PTilt, Time Dist. & Personal

By examining the selected features, it can be concluded that;

- Time-distance and Personal features have significant information on evaluating the grade of OA.
- As the severity of the illness increase, pelvic related features also hold discriminatory information.
- Altough comparisons of features of binary data sets are different, the combination of selected features are consistent with features used in Şen's study.
- Four features knee related, three features hip related, three features pelvic related and two features foot related.

By examining classification results, it can be concluded that;

- Set of binary classification of gait data produce better results than a single multiclass data set. Şen also reached the same conclusion [13].
- Control group can easily be separated from the patient group by using the methods implemented in this study. In other words, normal and abnormal subjects can be classified with %100 accuracy by the methods used in this study.
- Level 1 group has the lowest classification performance that supports criticism in medical domain about the inconsistencies of Grade 1.
- The classification performance of raw gait data set does not satisfactorily separate any level of OA. Therefore, dimension reduction adds significant contribution in analyzing gait data.

As mentioned before, there are many challenges on collecting gait data such as deviations of marker position, effects of age and body, trial-to trial differences and evaluator bias. Additionally, there is no agreed standard to merge and compare the result of different gait laboratories [6]. By considering the obstacles of acquiring reliable gait data, it can be concluded that the methods implemented in this study can help medical expert on evaluating grade of OA

6.2 Future Work and Discussion

Analysis of original and downsampled waveforms revealed that uniform downsampling might lose valuable information of the waveform. Instead of uniform downsampling, variable downsampling techniques might be implemented according to behavior of the gait cycle. For example, since there is no interaction with the ground during the swing phase, there is no need to sample swing phase of cycle of features calculated with GRF. Additionally, downsampling rate might be decreased during the peaks and slopes.

This study confirmed the importance of dimension reduction on analyzing gait data. SFS, PCA and Fourier analysis methods have significant contribution on reducing gait data. A further study on feature selection methods and implementations of these methods might increase the performance of the classifiers. Using with or without a feature selection method has significant difference on the performance of the classifier.

One of the main obstacles of classification an analyzing gait data is not having large number of samples comparing to feature space. Since gait analysis measurement specifications change system to system and trial to trial, stricter standards may help to scale measurements and data sets of different systems and different laboratories may be combined to get better classification accuracy.

REFERENCES

[1] Bertani A, Cappello A, Benedetti MG, Simoncini L, Catani F. “Flat foot functional evaluation using pattern recognition of ground reaction data”. Clin Biomech 14:484–93, 1999.

[2] Gitter A, McAnelly R. “The value of information resuling from instrumented gait analysis: The physiatrist.” In: DeLisa JA (ed.), Gait Analysis in the Science of Rehabilitation. Monograph 002. Baltimore: Department of Veteran Affairs, 69–75, 1998.

[3] Whittle M.W. “An Introduction to Gait Analysis”. Fourth Edition, Butterworth-Heinemann-Elsevier, 2007

[4] Hausdorff J.M., Peng C.K., Ladin Z., Wei J. Y., Goldberger A. L. “Is walking a random walk? Evidence for long-range correlations in stride interval of human gait” J. Appl. Physics, 78:349-358, 1995

[5] Yavuzer G. “Kinetic and kinematic characteristics of gait in patients with medial Knee arthritis”, Acta Orthop Scand; 73 (6):647–652, 2002

[6] Yavuzer G. “Three-dimensional quantitative gait analysis”, Acta Orthop Traumatol Turc; 43(2):94-101, 2009

[7] Yavuzer G., Sonel B., Kutlay F., Ergin S. “Use of Gait Analysis in the Treatment Decision-Making Process of Patients with Spastic Cerebral Palsy”, Journal of Physical Medicine and Rehabilitation, 51:1-5, 2005.

[8] Davis, M.A., Ettinger, W.H., et al., “Knee osteoarthritis and physical functioning: evidence from the NHANES I epidemiologic followup study” Journal of Rheumotology, 18:591–598, 1991

[9] U.S. National Library of Medicine National Institutes of Health http://www.nlm.nih.gov/medlineplus/ency/article/000423.htm, Accessed on December 2011

[10] Hochberg MC, Altman R, Brandt K, Clark B, Dieppe P, Griffin M, et al. “Guidelines for the medical management of osteoarthritis: part II. Osteoarthritis of the knee”, Arthritis & Rheumatism,Volume 38, Issue 11:1535–1540, 1995

[11] Kaufman K.R., Kenton R., Hughes C., Bernard F. Morrey B.F., Morrey M., An K., “Gait characteristics of patients with knee osteoarthritis”, Journal of Biomechanics, vol:34:907–915, 2001

[12] Kellgren JH, Lawrance JS “Radiological assessment of osteoarthritis**”**, Ann Rheum. Dis., vol. 16:494-501, 1957

[13] Şen N. “OAGAIT: A Decision Support System for Grading Knee Osteoarthritis Using Gait Data”, The Department of Information Science, Middle East Technical University, PhD Thesis, 2008

[14] Egton Medical Information Systems Limited
http://www.patient.co.uk/health/Osteoarthritis.htm ,
Accessed on December 2011

[15] Patton J., Sensory Motor Performance ProgramRehabilitation Institute of Chicago
http://www.smpp.northwestern.edu/~jim/kinesiology/partB_GaitMechanics.ppt.pdf
Accessed on December 2011

[16] Christopher J.C. Burges C.J. “Geometric Methods for Feature Extraction and Dimensional Reduction**”,** Kluwer Academic Publishers, 2004

[17] Cunningham P., “Dimension Reduction”, University College, Dublin, Technical Report UCD-CSI-2007-7, 2007

[18] Bellman R., “Adaptive Control Processes: A Guided Tour”, Princeton University Press, Princeton, 1961.

[19] Aksoy S. “Feature Reduction and Selection”
http://www.cs.bilkent.edu.tr/~saksoy/courses/cs551-Spring2008/slides/cs551_dimensionality.pdf
Accessed on December 2011

[20] Alpaydın E., “Introduction to Machine Learning”, MIT Press, 2004

[21] Saeys Y., Inza I., Larranaga P. “A review of feature selection techniques in Bioinformatics”, Bioinformatics, 23 (19): 2507-2517, 2007.

[22]. Guyon I., Elisse3 A.”An Introduction to Variable and Feature Selection”, Journal of Machine Learning Research, 3:1157-1182, 2003

[23] K. Fukunaga. “Introduction to Statistical Pattern Recognition”, Academic Press Professional Inc., San Diego, 1990.

[24] Carreira-Perpinan M. A.” A Review of Dimension Reduction Techniques ”, Technical Report CS-96-09, Dept. of Computer Science University of Sheffield, 1997

[25] Maaten L.J.P. van der, Postma E.O., Herik H.J. van den “Dimensionality Reduction: A Comparative review**”,** Tilburg Centre for Creative Computing, Tilburg University, 2009

[26] Osuna R. G.,”Introduction to Pattern Analysis”, Texas A&M University http://research.cs.tamu.edu/prism/lectures.htm,
Accessed on December 2011

[27] Fodor I. K., “A survey of dimension reduction techniques**”,** Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, 2002

[28] Chau T. “A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods”, Gait and Posture, 13:49–66, 2001

[29] Deluzio K.J., Astephen J.L., “Biomechanical features of gait waveform data associated with knee osteoarthritis an application of principal component analysis”, Gait & Posture, 25:86–93,2007

[30] O’Malley MJ, Abel MF, Damiano DL, Vaughan CL. “Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters”. IEEE Trans Rehabil Eng Dec, 5(4):300–9, 1997.

[31] Tan T, Guan L, Burne J. “A real-time image analysis system for computer-assisted diagnosis of neurological disorders”, Real-time Imaging, (4):253–69, 1999.

[32] Merkle LA, Layne CS, Bloomberg JJ, Zhang JJ. “Using factor analysis to identify neuromuscular synergies during treadmill Walking”, Journal of Neuroscience Methods, 82:207–14, 1998.

[33] Greenacre M, Blasius J. “Correspondence Analysis in the Social Sciences: Recent Developments and Applications”, London, Academic Press, 1994.

[34] Loslever P, Laassel E, Angue JC. “Combined statistical study of joint angles and ground reaction forces using component and multiple correspondence-analysis”, IEEE Trans. Biomed. Eng, 41(12):1160–7, 1994.

[35] Marghitu DB, Nalluri P. “An analysis of greyhound gait using Wavelets”, Journal of Electromyography and Kinesiology, 7(3):203–12, 1997.

[36] Tamura T, Sekine M, Ogawa M, Togawa T, Fukui Y., “Classification of acceleration waveforms during walking by wavelet transform”, Methods of Information in Medicine, 36:356–9, 1997.

[37] Chau T. “A review of analytical techniques for gait data. Part 2: neural network and wavelet methods”, Gait and Posture, 13:102–120, 2001

[38] Wu J., Wang J., Liu l., ”Feature extraction via KPCA for classification of gait patterns”, Human Movement Science, 26:393–411, 2007

[39] Haykin, S. , “Neural networks: A comprehensive foundation (2nd ed.)”, New York, Prentice-Hall., 1999

[40] Scholkopf B., Smola, A. J., & Muller, K.-R. “Nonlinear component analysis as a kernel eigenvalue Problem”, Journal of Neural Computation, 10:1299–1319, 1998

[41] Kim, K. I., Franz, M. O., & Scholkopf, B. “ Iterative kernel principal component analysis for image modeling”., IEEE Transactions on Pattern Analysis and Machine Intelligence, 27:1352–1366, 2005.

[42] Deluzio K.J., Wyss U.P., Zee B., Costigan P. A., "Principal Component Models of Knee Kinematics and Kinetics: Normal vs. pathological Gait Patterns", Human Movement Science, 16:201-207, 1997

[43] Jolliffe I.T., ”Principal Component Analysis”, Springer-Verlag, 2002

[44] The Fourier Transforms
http://www.thefouriertransform.com/transform/fourier.php#introduction
Accessed on December 2011

[45] Sethares W.A. “Rhythm and Transforms”, Springer-Verlag London Limited, 2007

[46] Park T.H. ,“Introduction to Digital Signal Processing”, World Scientific Publishing, 2010

[47] Tan L. “Digital Signal Processing Fundamentals and Applications”, Elsevier, 2008

[48] VISOL, 3D Motion Analysis System
http://www.kwon3d.com/theory/filtering/lpass.html,
Accessed on December 2011

[49] Devijver P., Kittler J., “Pattern recognition: A statistical approach”, Prentice Hall, Englewood Cliffs, London, 1982

[50] Cao L.J. et al “A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine”, Neurocomputing, 55:321 – 336, 2003

[51] Müller K.R. , Mika S. , Ratsch G. , Tsuda K. , Scholkopf B. , “An Introduction to Kernel Based Learning Algorithms”, IEEE Transactions On Neural Networks, Vol. 12 No:2, 2001

[52] Vapnik V.N., "Statistical Learning Theory", Wiley, New York, 1998

[53] Duda R.O., Hart P.E., Stork D.G., “Pattern Classification”, John Wiley and Sons, New York, 2001

[54] Rote G. “Computing the minimum Hausdorff distance between two point sets on a line under translation.” Information Processing Letters, 38:123-127, 1991

[55] Huttenlocher D.P., Klanderman G.A., Rudgelidge W. J. "Comparing the images using the Hausdorff distance", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 15 No:9:850-863, 1993

[56] Matlab R2009a, The MathWorks Inc., USA

[57] RapidMiner 5.3.008, Copyright 2001-2013 by Rapid-I GmbH and contributors

[58] Vicon Plug- In Gait Manuel Version 1.0
Received via e-mail from Vicon Support Services for academical study purposes. http://www.vicon.com/support/
Accessed on December 2011

[59] Hart D.J., Spector T.D., "Kellgren & Lawrence grade 1 osteophytes in the knee—doubtful or definite?", Osteoarthritis and Cartilage, 11:149–150
OsteoArthritis Research Society International, 2003

[60]. Massicotte F." Epidemiology of osteoarthritis", Osteoarthritis Research Unit, University of Montreal Hospital Research Centre, Understanding Osteoarthritis from Bench to Bedside, 2011: 1-26

[61] Subashini T.S., Ramalingam V., Palanivel S., "Breast mass classification based on cytological patterns using RBFNN and SVM", Expert Systems with Applications 36:5284–5290, 2009

[62] ADAM Multimedia Encyclopedia
http://adam.about.net/reports/Osteoarthritis.htm,
Accessed on December 2011

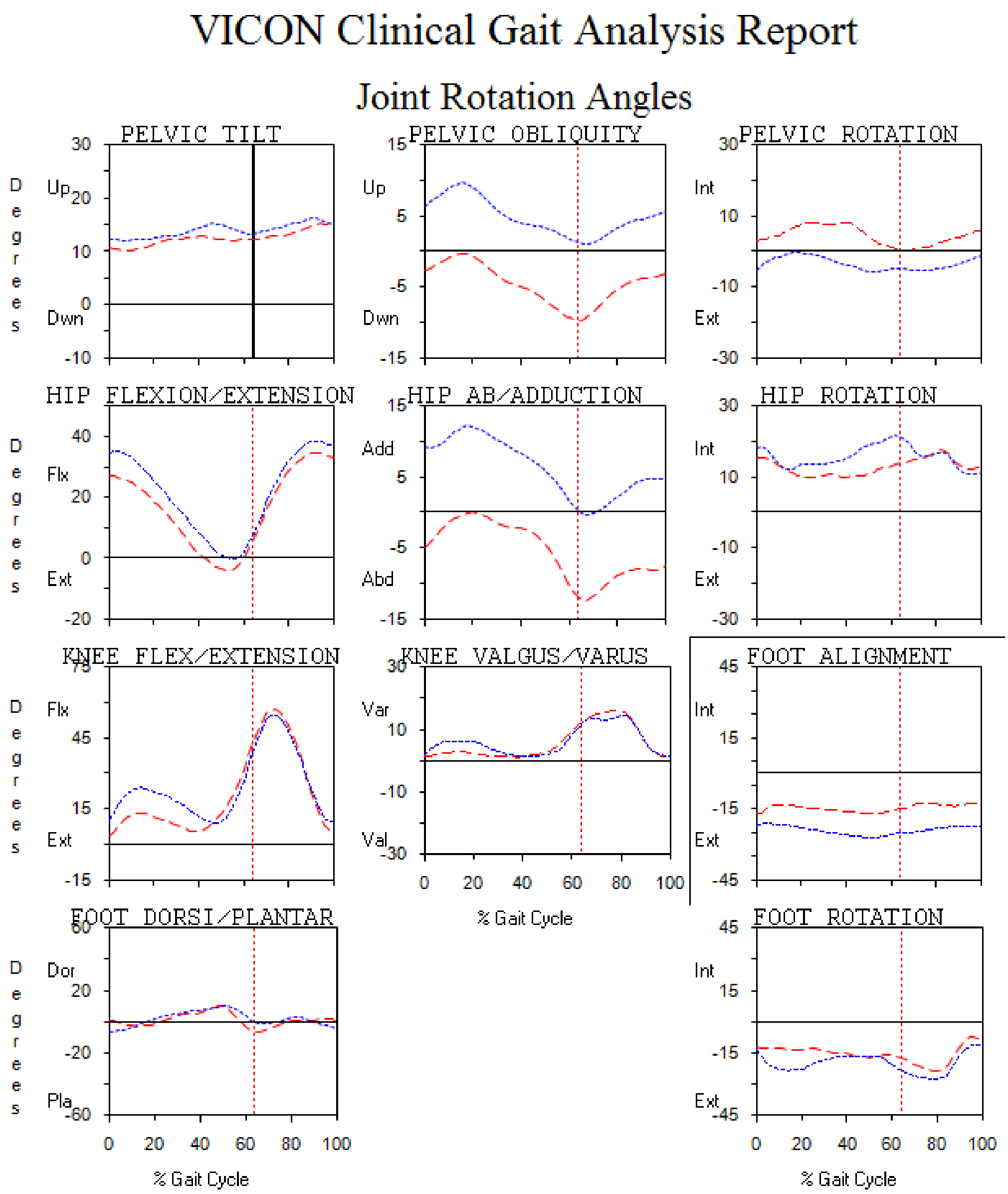
[63] Chih-Wei Hsu, Chih-Jen Lin "A Comparison of Methods for Multiclass Support Vector Machines" IEEE Transactions on Neural Networks, Vol. 13, No: 2,:415, 2002

[64] Mikel Galar, AlbertoFernandez, Edurne Barrenechea, Humberto Bustince, Francisco Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes", Pattern Recognition, 44:1761–1776, 2011

[65] Evrim Itır Karaç, “Model Selection For Multi Class Support Vector Machines”, Boğaziçi University, MSc Thesis, 2000

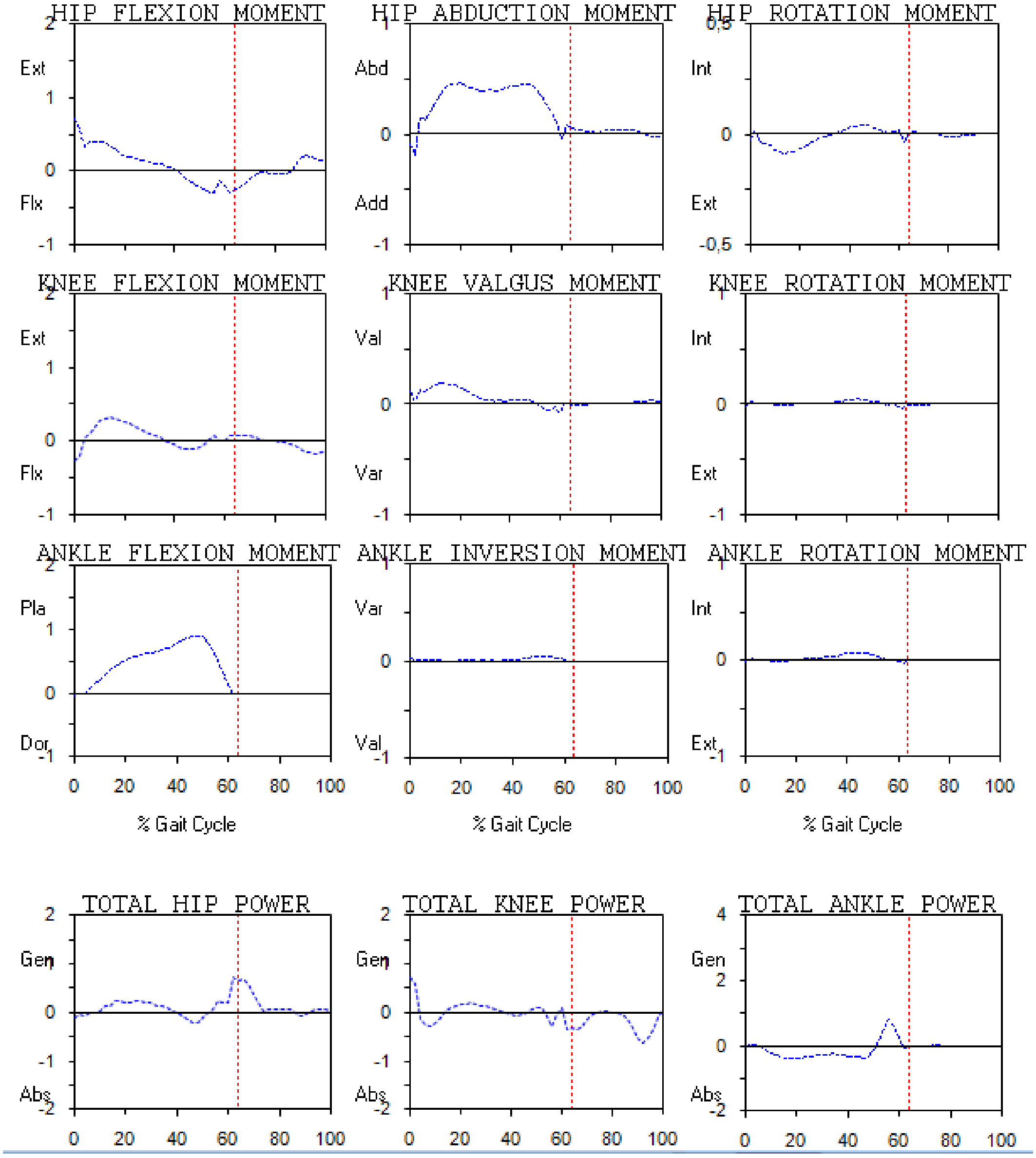
APPENDICES

APPENDIX A A SAMPLE OF VICON CLINICAL GAIT ANALYSIS REPORT



VICON Clinical Gait Analysis Report

Joint Net Moments and Powers



APPENDIX B OUTLIER ANALYSIS STATISTICS & MATLAB CODE

Outlier Analysis Results

Feature	Level	Total Number of Subjects	Total Number of Subjects Without Outliers	% of Subjects Without Outliers
HAbd	0	46	45	97
	1	20	20	100
	2	69	64	92
	3	45	41	91
FRot	0	46	43	93
	1	20	20	100
	2	69	65	94
	3	45	45	100
KFlex	0	46	45	97
	1	20	20	100
	2	69	67	97
	3	45	42	93
HRot	0	46	45	97
	1	20	20	100
	2	69	65	94
	3	45	44	97
FDor	0	46	44	95
	1	20	19	95
	2	69	66	95
	3	45	45	100
KMRot	0	46	44	95
	1	20	19	95
	2	69	66	95
	3	45	42	93
KMVal	0	46	46	100
	1	20	19	95
	2	69	66	95
	3	45	43	95
FMDor	0	46	46	100
	1	20	19	95
	2	69	63	91

	3	45	43	95
KPTot	0	46	46	100
	1	20	20	100
	2	69	66	95
	3	45	43	95
KPFlex	0	46	45	97
	1	20	20	100
	2	69	66	95
	3	45	44	97
KPVal	0	46	45	97
	1	20	19	95
	2	69	66	95
	3	45	42	93
KPRot	0	46	45	97
	1	20	19	95
	2	69	65	94
	3	45	44	97
KRot	0	46	45	97
	1	20	19	95
	2	69	65	94
	3	45	44	97
HFlex	0	46	44	95
	1	20	20	100
	2	69	64	92
	3	45	44	97
PObliq	0	46	44	95
	1	20	20	100
	2	69	63	91
	3	45	42	3
PTilt	0	46	43	93
	1	20	20	100
	2	69	65	94
	3	45	46	100
PRot	0	46	43	93
	1	20	19	95
	2	69	65	94
	3	45	41	91
Fpro	0	46	44	95
	1	20	20	100
	2	69	66	95
	3	45	44	97
KVal	0	46	44	95
	1	20	19	95

	2	69	65	94
	3	45	42	93
HMRot	0	46	44	95
	1	20	18	90
	2	69	64	92
	3	45	40	88
HMAbd	0	46	46	100
	1	20	18	90
	2	69	66	95
	3	45	43	95
HMFlex	0	46	44	95
	1	20	20	100
	2	69	67	97
	3	45	43	95
KMFlex	0	46	46	100
	1	20	19	95
	2	69	66	95
	3	45	41	91
FMAbd	0	46	46	100
	1	20	19	95
	2	69	65	94
	3	45	41	91
FMRot	0	46	44	95
	1	20	18	90
	2	69	66	95
	3	45	44	97
HPTot	0	46	46	100
	1	20	19	95
	2	69	67	97
	3	45	43	95
HPFlex	0	46	44	95
	1	20	20	100
	2	69	67	97
	3	45	42	93
HPAbd	0	46	44	95
	1	20	19	95
	2	69	67	97
	3	45	44	97
HPRot	0	46	44	95
	1	20	19	95
	2	69	65	94
	3	45	43	95
APTot	0	46	46	100

	1	20	19	95
	2	69	66	95
	3	45	42	93
APDor	0	46	46	100
	1	20	20	100
	2	69	66	95
	3	45	43	95
APAbd	0	46	45	97
	1	20	18	90
	2	69	64	92
	3	45	41	91
APRot	0	46	45	97
	1	20	19	95
	2	69	66	95
	3	45	44	97

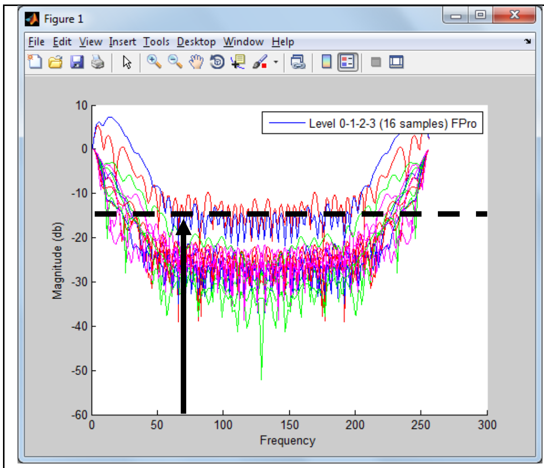
Matlab Function For Outlier Analysis

```
function Outlier(M,level,sheet)
% Sample call for the function
%FRot0=xlsread('C:\Data\gradeL0.xls','FRot');
% Outlier(FRot0,0,’Frot’)
sM=std(M);
sz=size(M);
mn=mean(M);
tempM=zeros(sz(1),sz(2));
str='';
r=1;
t=0;
for i=1:sz(1)
    outlier=0;
    outsayi=sz(2)*0.25;
    for j=1:sz(2)
       if (M(i,j)>mn(j)+2*sM(j)) || (M(i,j)<mn(j)-2*sM(j))
           outlier=outlier+1;
       end
     end
     if outlier<outsayi
        yeniM(r,:)=M(i,:);
        r=r+1;
        tempM(i,:)=M(i,:);
     else
         t=t+1;
         outlierCount(t)=i;
         str=strcat(str,',', num2str(i));
    end
end
mn2=mean(yeniM);
if t>0
    for k=1:t
        tempM(outlierCount(k),:)=mn2;
    end
end
xlswrite(strcat('C:\Code\SgradeL75PRCX',level,'.xls'),tempM,sheet);
```

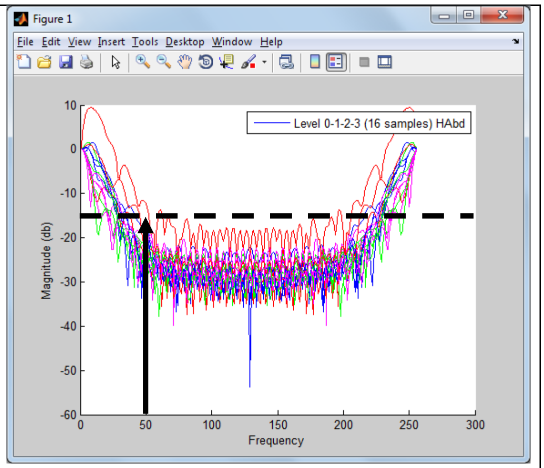
```
fid=fopen('C:\Code\yenikayit75PRC.txt','a');
fprintf(fid,'%s level:%s Kayıt:%d KayıtWO: %d  Yüzde: %d
%s\r\n',sheet,level,sz(1),r,(r/sz(1))*100,str);
fclose(fid);
end
```

APPENDIX C FOURIER ANALYSIS & DOWNSAMPLING RATE DETERMINATON & MATLAB CODE

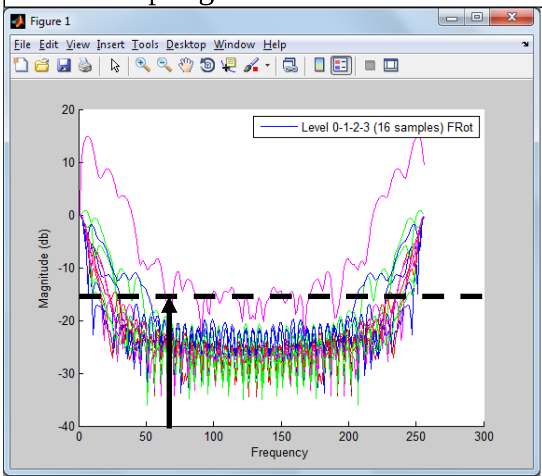
To determine the downsampling rate of a feature, among all subjects largest bandwidth (BW) at -15 db level was selected. If BW was less than or equal to 64 (half of Nyquist rate), downsampling rate was determined as “2”. If BW was greater than 64, no downsampling was performed.



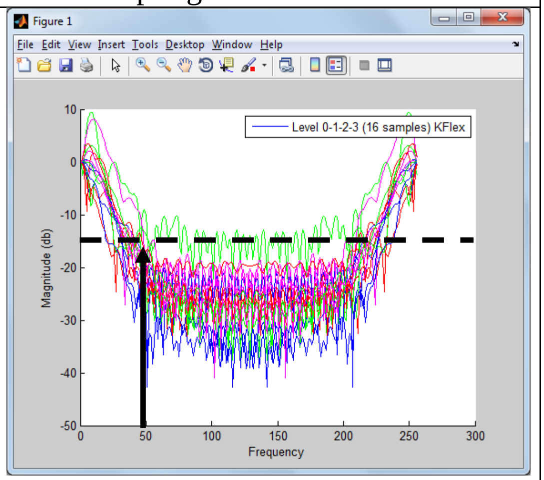
Downsampling=2



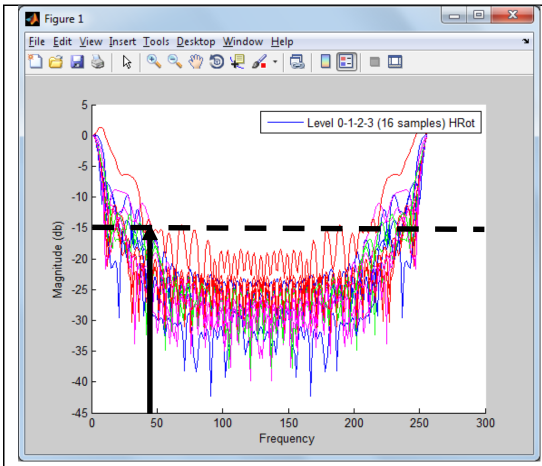
Downsampling=2



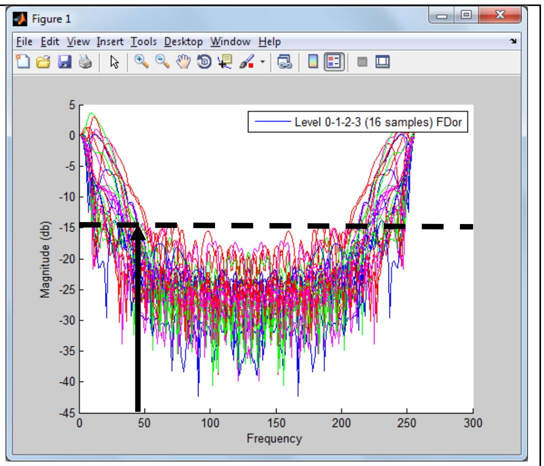
Downsampling=2



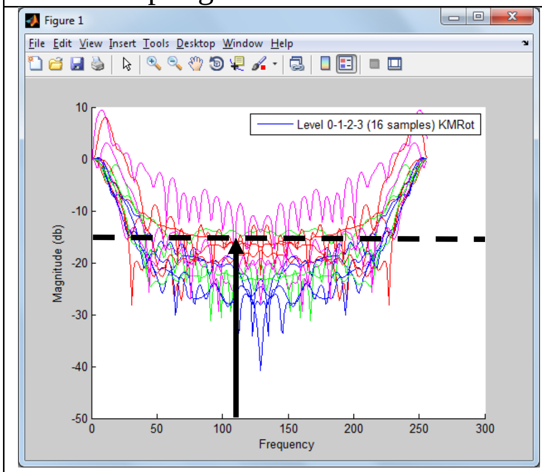
Downsampling=2



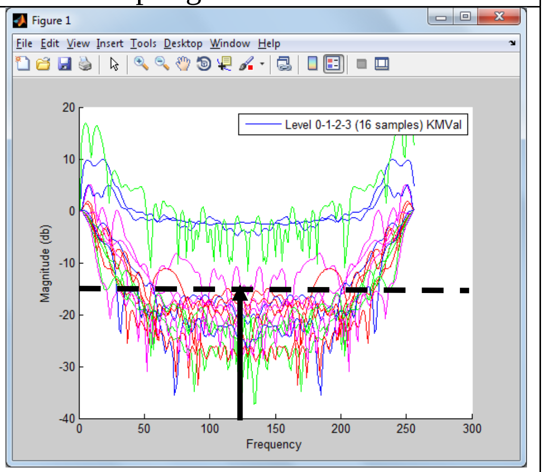
Downsampling=2



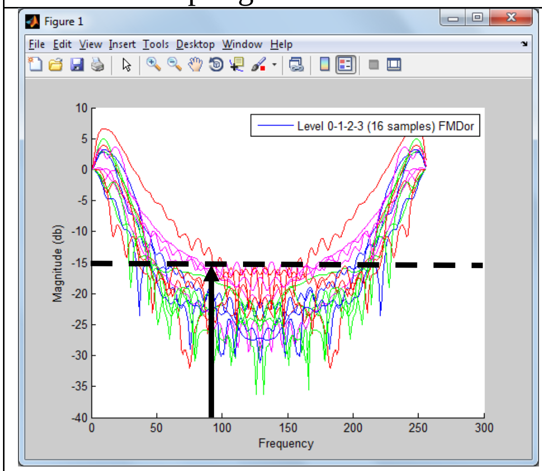
Downsampling=2



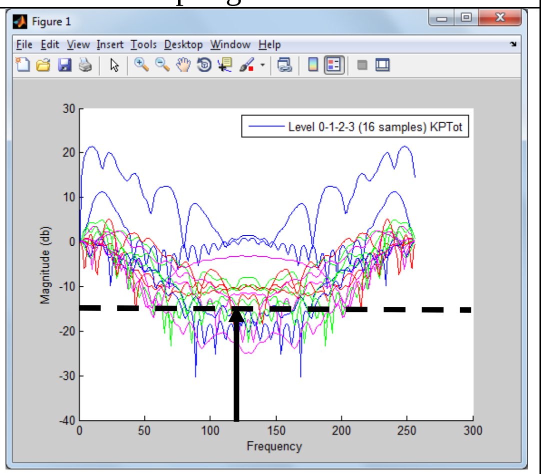
No downsampling



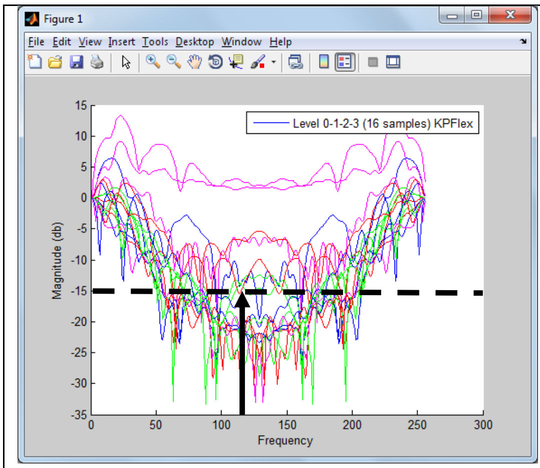
No downsampling



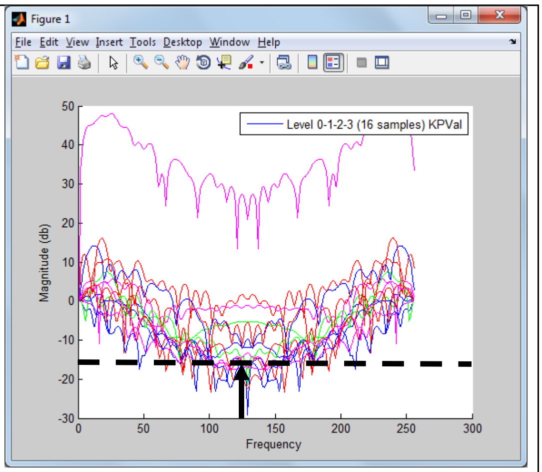
No downsampling



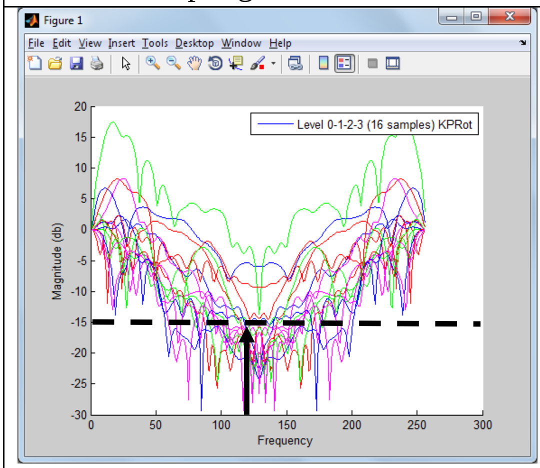
No downsampling



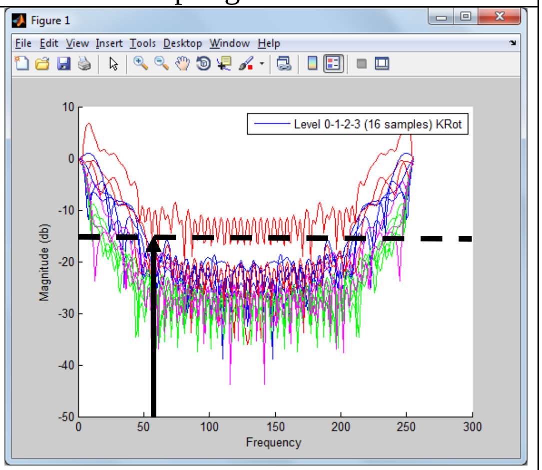
No downsampling



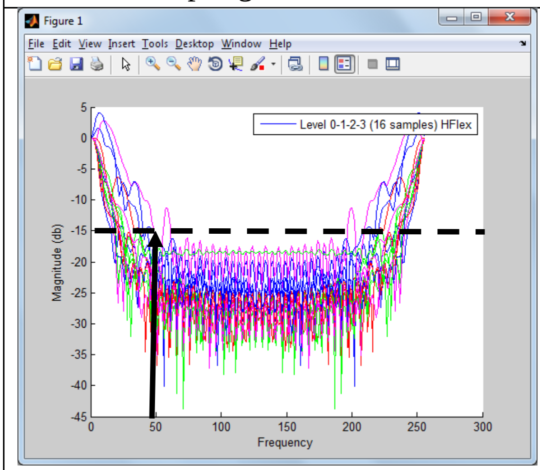
No downsampling



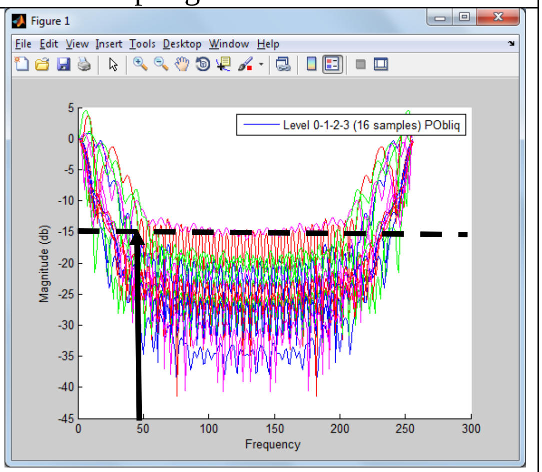
No downsampling



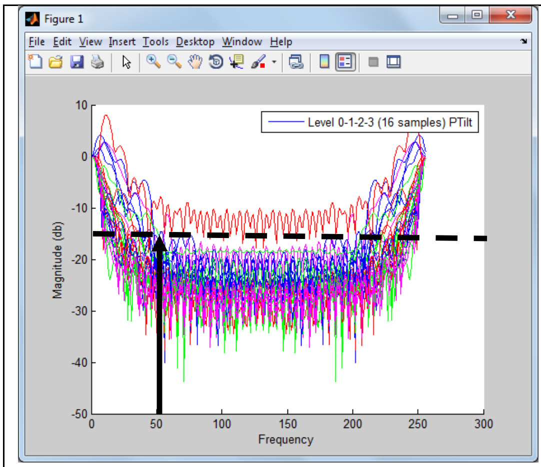
Downsampling=2



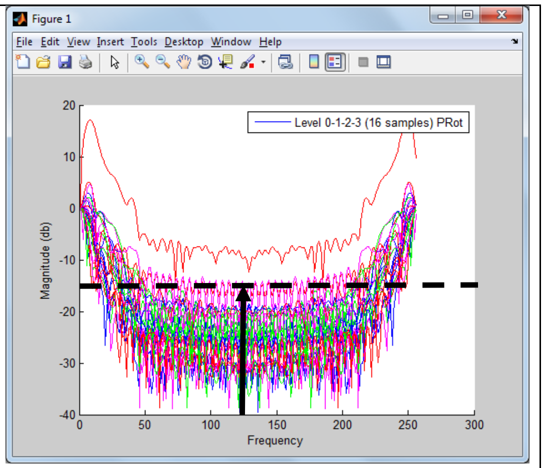
Downsampling=2



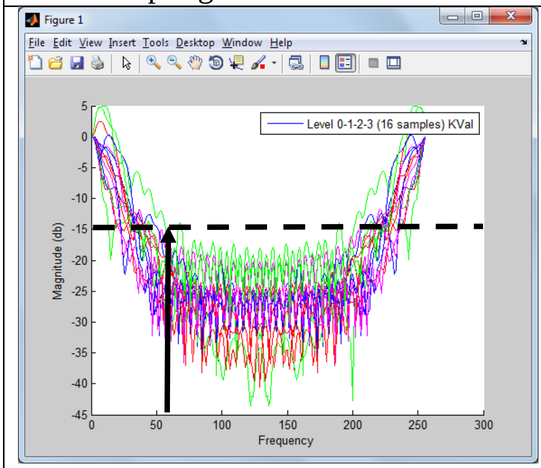
Downsampling=2



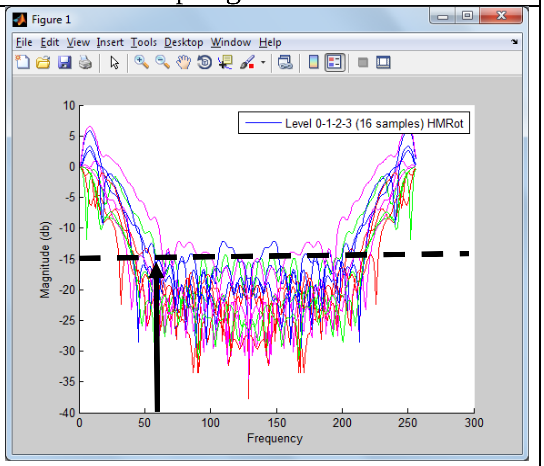
Downsampling=2



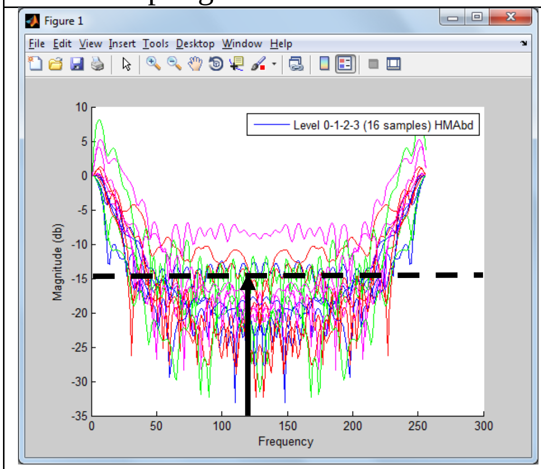
No downsampling



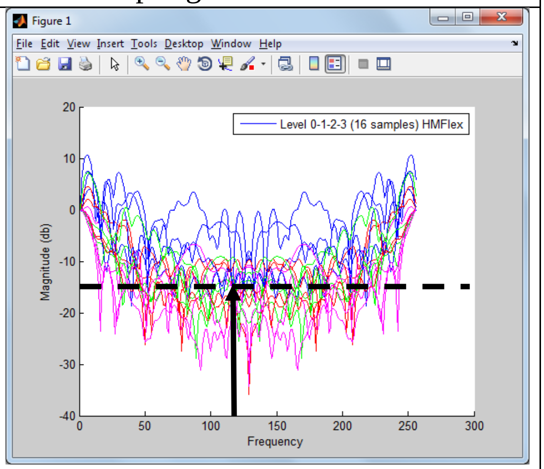
Downsampling=2



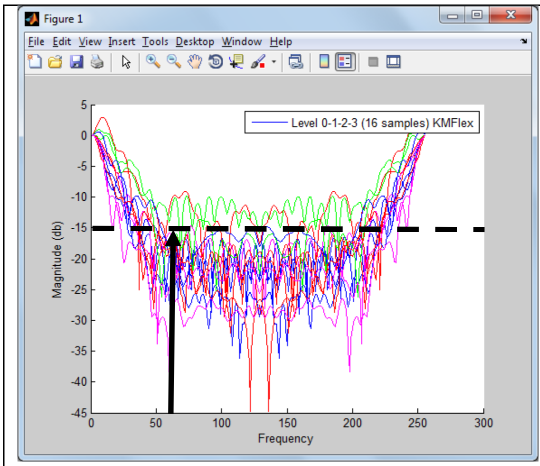
Downsampling=2



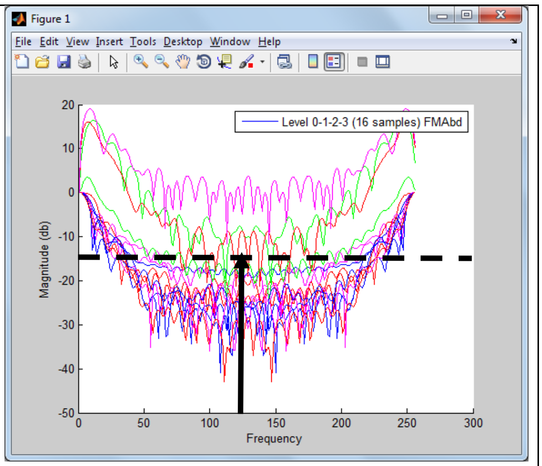
No downsampling



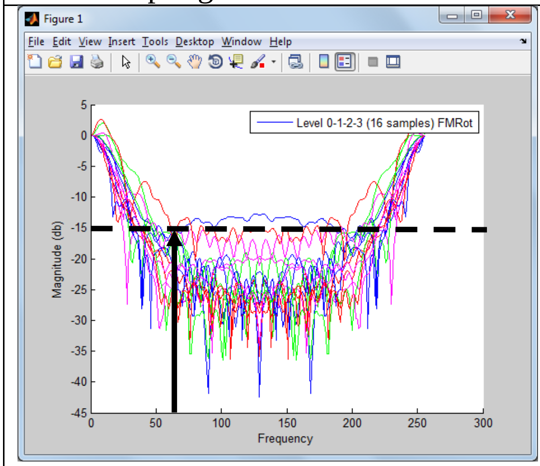
No downsampling



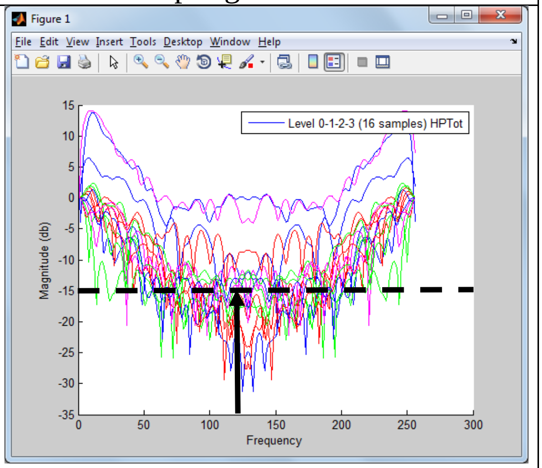
Downsampling=2



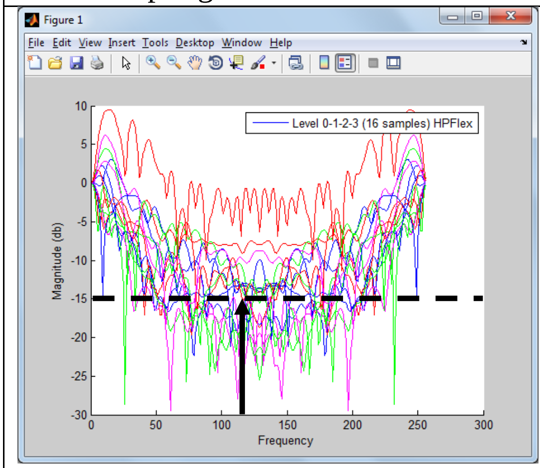
No downsampling



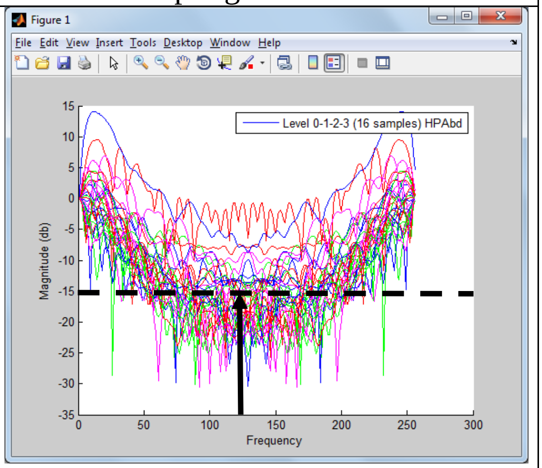
Downsampling=2



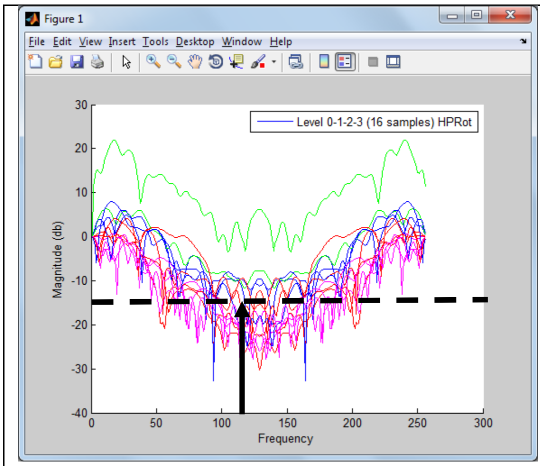
No downsampling



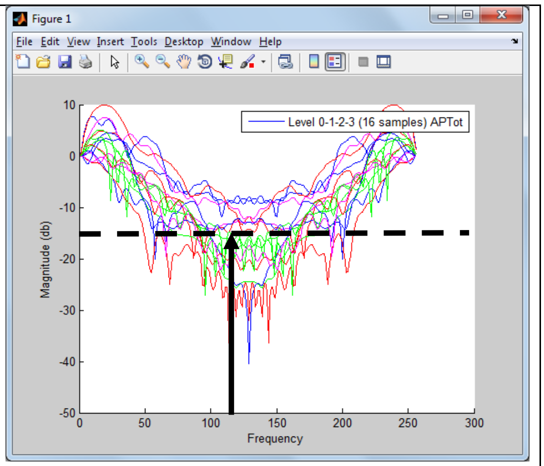
No downsampling



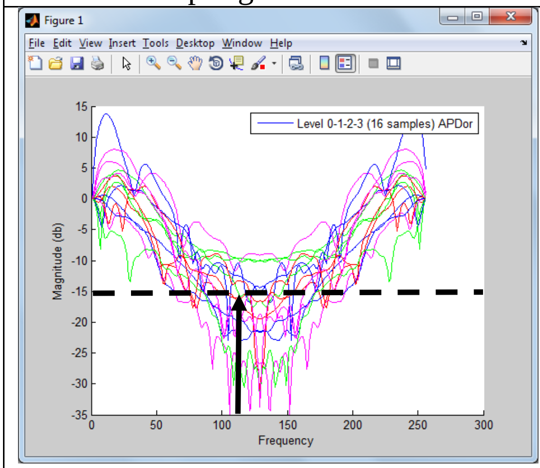
No downsampling



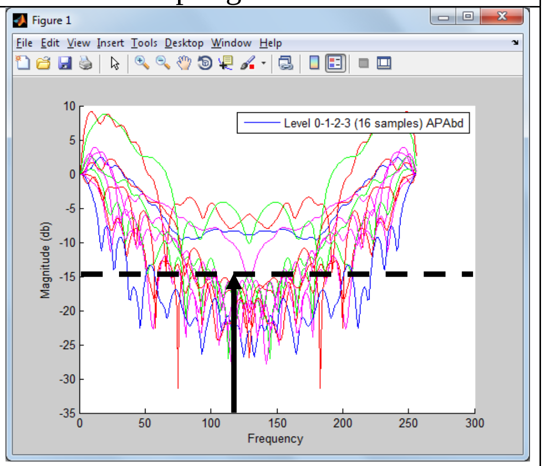
No downsampling



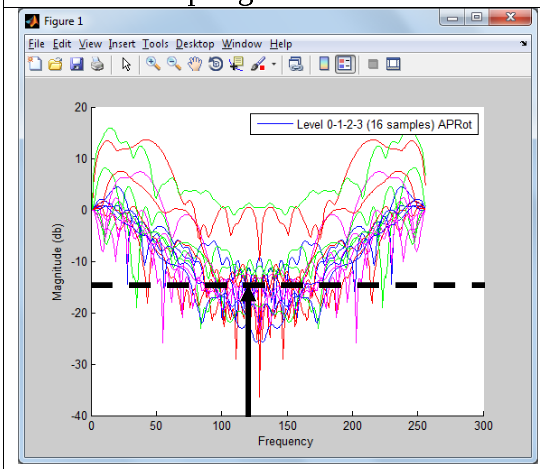
No downsampling



No downsampling

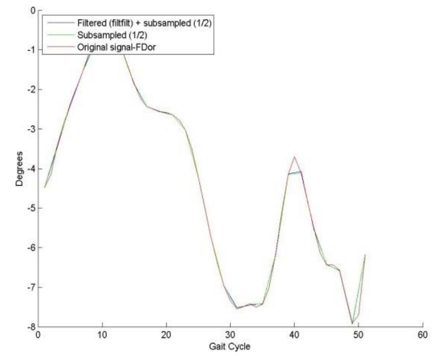


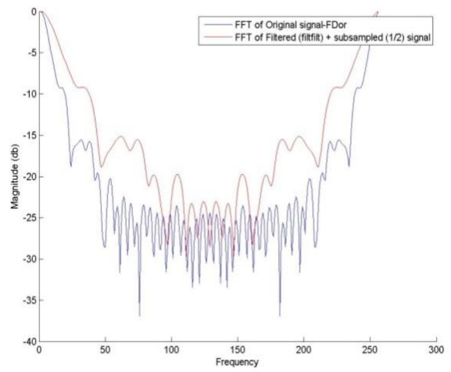
No downsampling

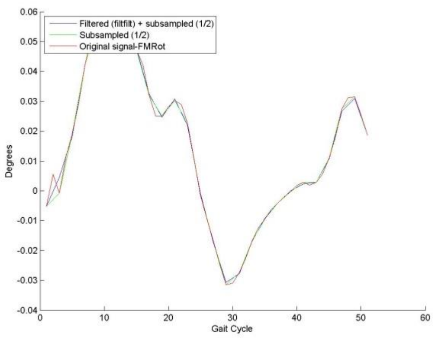


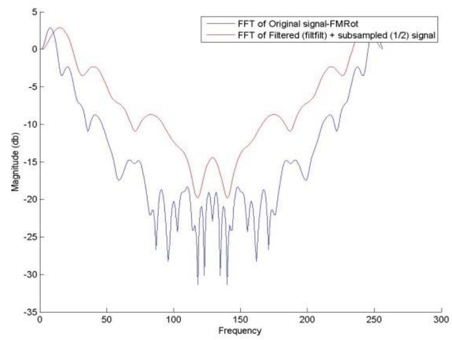
No downsampling

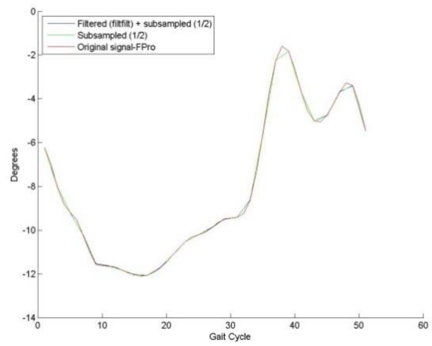
Original and Downsampled Signals

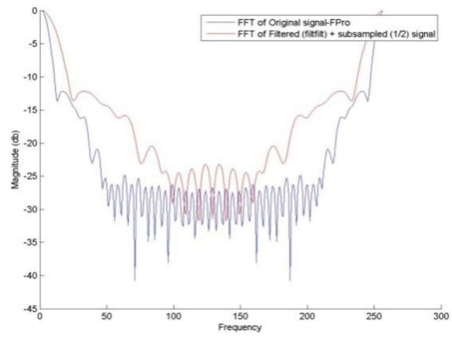


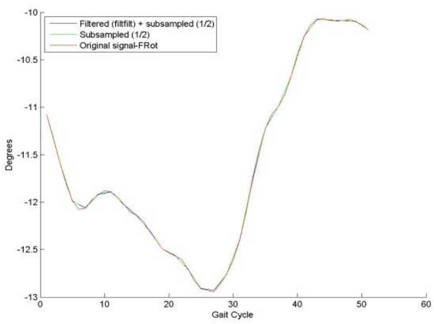


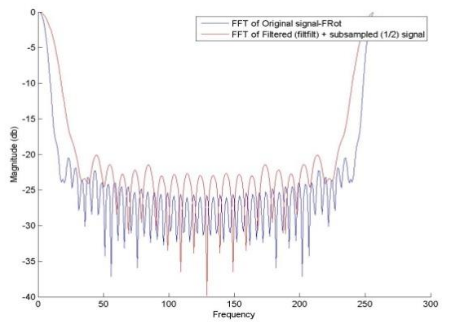


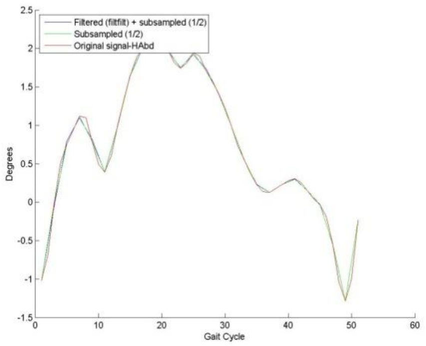


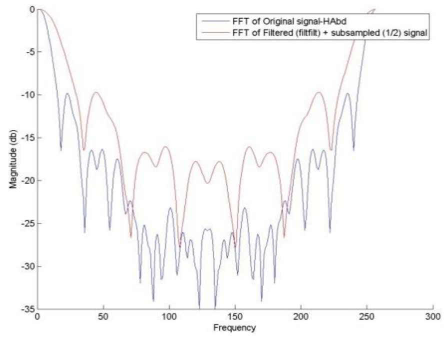


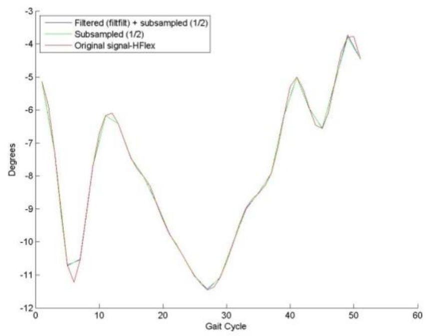


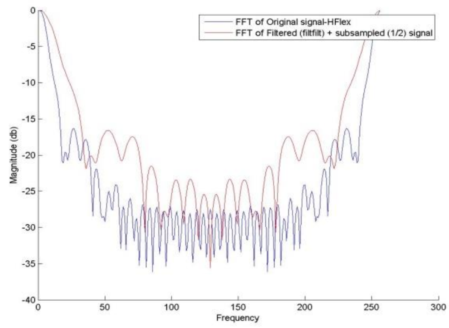


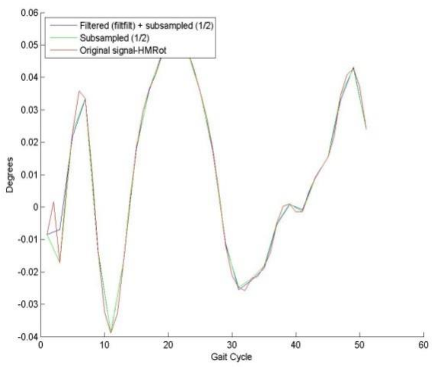


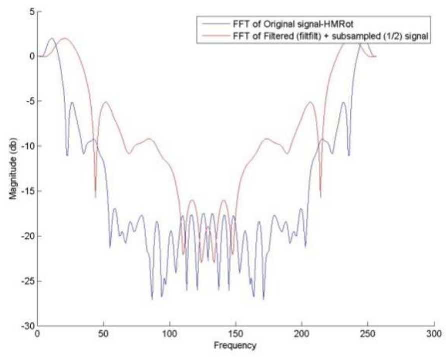

Filtered (filtfilt) + subsampled (1/2)
Subsampled (1/2)
Original signal-HAbd
Degrees
Gait Cycle

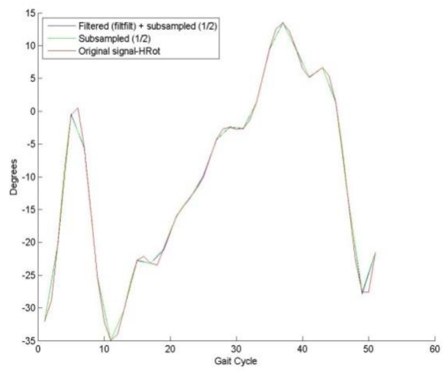

FFT of Original signal-HAbd
FFT of Filtered (filtfilt) + subsampled (1/2) signal
Magnitude (db)
Frequency

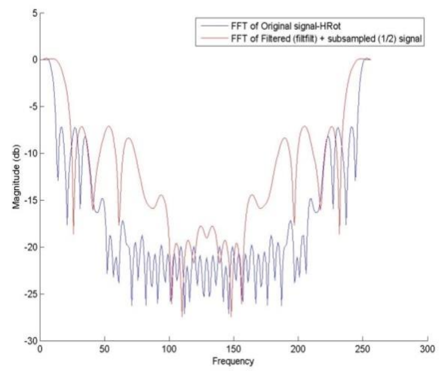

Filtered (filtfilt) + subsampled (1/2)
Subsampled (1/2)
Original signal-HFlex
Degrees
Gait Cycle

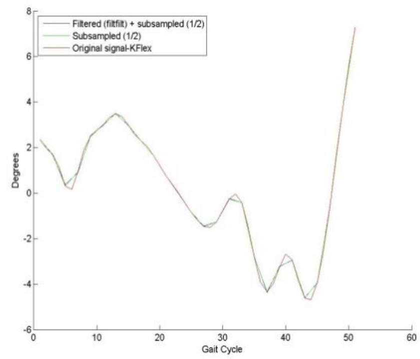

FFT of Original signal-HFlex
FFT of Filtered (filtfilt) + subsampled (1/2) signal
Magnitude (db)
Frequency

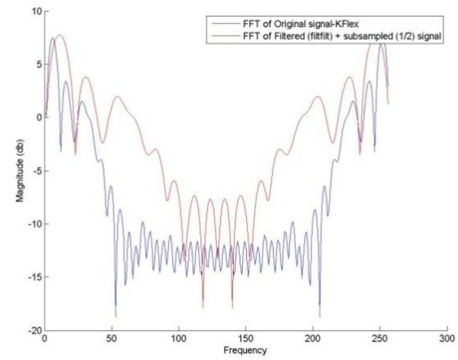

Filtered (filtfilt) + subsampled (1/2)
Subsampled (1/2)
Original signal-HMRot
Degrees
Gait Cycle

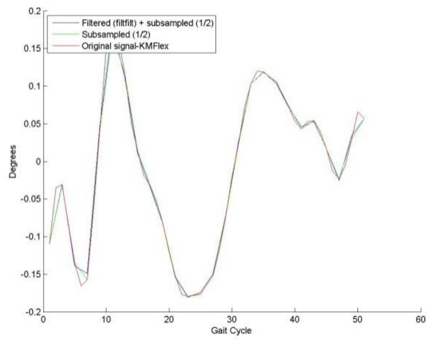

FFT of Original signal-HMRot
FFT of Filtered (filtfilt) + subsampled (1/2) signal
Magnitude (db)
Frequency

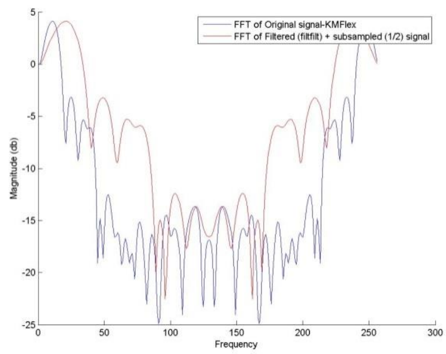

Filtered (filtfilt) + subsampled (1/2)
Subsampled (1/2)
Original signal-HRot
Degrees
Gait Cycle

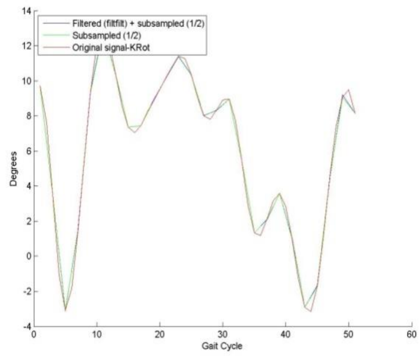

FFT of Original signal-HRot
FFT of Filtered (filtfilt) + subsampled (1/2) signal
Magnitude (db)
Frequency

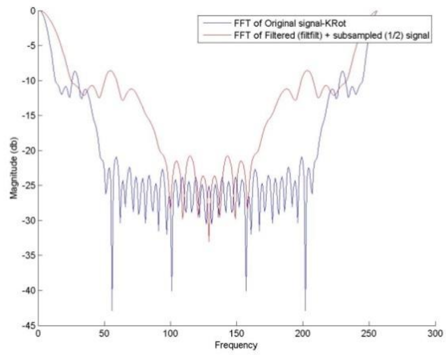

Filtered (filtfilt) + subsampled (1/2)
Subsampled (1/2)
Original signal-KFlex
Degrees
Gait Cycle

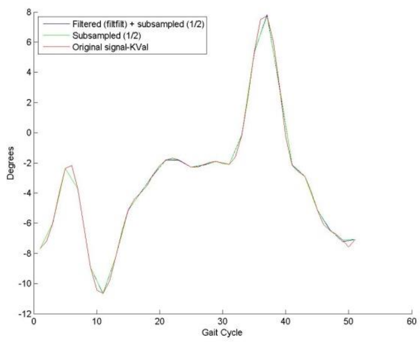

FFT of Original signal-KFlex
FFT of Filtered (filtfilt) + subsampled (1/2) signal
Magnitude (db)
Frequency

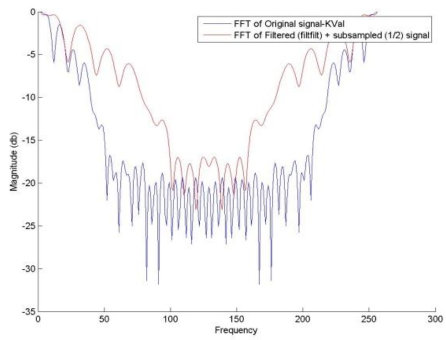

Filtered (filtfilt) + subsampled (1/2)
Subsampled (1/2)
Original signal-KMFlex
Degrees
Gait Cycle

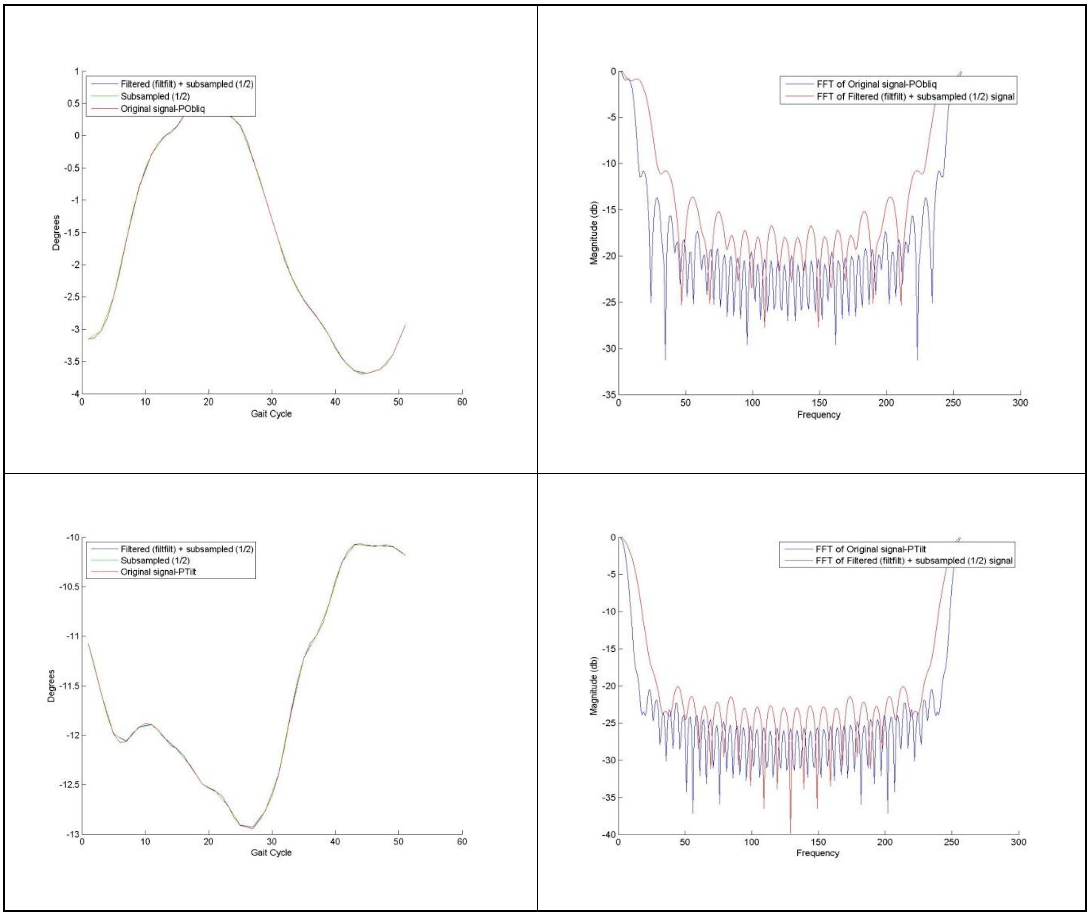

FFT of Original signal-KMFlex
FFT of Filtered (filtfilt) + subsampled (1/2) signal
Magnitude (db)
Frequency


Filtered (filtfilt) + subsampled (1/2)
Subsampled (1/2)
Original signal-KRot
Degrees
Gait Cycle


FFT of Original signal-KRot
FFT of Filtered (filtfilt) + subsampled (1/2) signal
Magnitude (db)
Frequency


Filtered (filtfilt) + subsampled (1/2)
Subsampled (1/2)
Original signal-KVal
Degrees
Gait Cycle


FFT of Original signal-KVal
FFT of Filtered (filtfilt) + subsampled (1/2) signal
Magnitude (db)
Frequency



14 features listed above were downsampled by 2. As it can be seen graphs on the left downsampled signal almost matches the original signal. There are some differences on sharp slopes and peeks.

Matlab Code for Fourier Analysis

```
HMRot0=xlsread('C:\Data\gradeL0.xls','HMRot');

[b1,a1]=butter(4,0.500,'low');

figure(2);
z=mean(HMRot0); z=z'.*hamming(51);
y=fft(z,256);
y=y/abs(y(1)); y=10*log10(abs(y));
plot(y,'b');

v1=zeros(51,1); v1(1)=1.0;
v=filtfilt(b1,a1,v1);
y=fft(v,256);
y=y/abs(y(1)); y=10*log10(abs(y));
legend('Impulse function (Filtfilt)');
sz=size(HMRot0);
fl=zeros(sz(1),sz(2));
for i=1:sz(1)
    fl(i,:)=filtfilt(b1,a1, HMRot0(i,:););
end
```

```
dF=downsample(fl',2);
dnF=downsample(HMRot0',2);
dnF=dnF';
dF=dF';
k=[1,3,5,7,9,11,13,15,17,19,21,23,25,27,29,31,33,35,37,39,41,43,45,4
7,49,51];
figure(4);
hold on
plot(k,dF(1,:),'b');
plot(k,dnF(1,:),'g');
plot(HMRot0(1,:),'r');
h=Legend('Filtered (filtfilt) + subsampled (1/2)','subsampled
(1/2)','Original signal HMRot');
set(h,'location','NorthWest');
figure(2);
hold on;
z=dF(1,:); z=z'.*hamming(26);
z=fft(z,256);
z=z/abs(z(1)); z=10*log10(abs(z));
plot(z,'r');
legend('Original signal fft HMRot','Filtered (filtfilt) + subsampled
(1/2) fft ');
xlswrite('C:\Data\gradeL0FFT.xls',dF,'HMRot');
```

APPENDIX D MATHEMATICAL DETAILS OF PCA AND SVM

Mathematical Derivation of PCA

N dimensional random vector x can be expressed with linear combination of orthonormal basis vectors $[\varphi_1|\varphi_2|\ldots|\varphi_N]$ such as

$$x = \sum_{i=1}^{N} \boldsymbol{y}_i \boldsymbol{\varphi}_i \ where\ \varphi_i \mid \varphi_i \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

If x is reexpressed with M basis vectors such thatM<N , pre-selected constants b_i should be added to the equation for the range of $[y_{M+1}, \ldots, y_N]^T$.

$$\hat{x}(M) = \sum_{i=1}^{M} \boldsymbol{y}_i \boldsymbol{\varphi}_i + \sum_{i=M+1}^{N} \boldsymbol{b}_i \boldsymbol{\varphi}_i$$

The representation error becomes

$$\Delta x(M) = x - \hat{x}(M) = \textstyle\sum_{i=1}^{N} \boldsymbol{y}_i \boldsymbol{\varphi}_i - \left(\sum_{i=1}^{M} \boldsymbol{y}_i \boldsymbol{\varphi}_i + \sum_{i=M+1}^{N} \boldsymbol{b}_i \boldsymbol{\varphi}_i\right)$$

$$= \textstyle\sum_{i=1}^{M} (y_i - b_i) \varphi_i$$

The error of the new linear combination can be calculated by the mean-squared magnitude of Δx.

The optimality problem becomes finding the basis vectors φ_i and constants b_i to minimize mean-square error.

$$\bar{\varepsilon}^2 \Delta(M) = E[\,[x(M)]^2] = E\left[\textstyle\sum_{i=M+1}^{N} \sum_{j=M+1}^{N} (y_i - b_i)(y_j\, b_j)\, \varphi_i^T \varphi_j\right] = \textstyle\sum_{i=M+1}^{N} E(y_i - b_i)^2$$

The optimality problem of finding b_i can be solved taking partial derivative of the objective function ($\bar{\varepsilon}^2 \Delta(M)$).

$$\frac{\partial}{\partial b_i} E[(y_i - b_i)^2] = -2(E[y_i] - b_i) = 0 \Rightarrow b_i = E[y_i]$$

The derivation function can be written as

$$\bar{\varepsilon}^2(M) = \sum_{i=M+1}^{N} E[(y_i - E[y_i])^2] = \sum_{i=M+1}^{N} E[(x\varphi_i - E[x\varphi_i])^T(x\varphi_i - E[x\varphi_i])]$$
$$= \sum_{i=M+1}^{N} \varphi_i^T E[(x - E[x])(x - e[x])^T]\varphi_i = \sum_{i=M+1}^{N} \varphi_i^T \Sigma_x \varphi_i$$

Since the basis vectors are orthonormal to each other the optimality function is subject to orthonormality constraint which is realized by incorporating Lagrange multipliers λ_i

$$\bar{\varepsilon}^2(M) = \sum_{i=M+1}^{N} \varphi_i^T \Sigma_x \varphi_i + \sum_{i=M+1}^{N} \lambda_i(1 - \varphi_i^T \varphi_i)$$

Taking partial derivative of the above function with respect to basis vectors, the solution can be expressed as eigenvectors and eigenvalues of the covariance matrix Σ_x.

$$\frac{\partial}{\partial\varphi_i}\bar{\varepsilon}^2(M) = \frac{\partial}{\partial\varphi_i}[\sum_{i=M+1}^{N} \varphi_i^T \sum_x \varphi_i + \sum_{i=M+1}^{N} \lambda_i(1 - \varphi_i^T\varphi_i)] = 2(\sum_x \varphi_i - \lambda_i\varphi_i) = 0$$

$$\Rightarrow \sum_x \varphi_i = \lambda_i \varphi_i$$

Support Vector Machine (SVM)

SVM is a machine learning technique developed by Cortes and Vapnik in 1995 and mainly used in supervised classification problems. SVM works by maping the feature space into higher dimensional feature space and constructing the classifying hyperplane on the higher space [61].

Linearly Separable Binary Classification

There are many solutions finding a separating hyperplane for a linearly separable dataset of two classes as seen in figure below.

$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, $x \in R^D$, $y \in \{-1, +1\}$

where vector x represents the coordinates of a sample in input space and y represents the class label which is +1 and -1 in binary classification case.

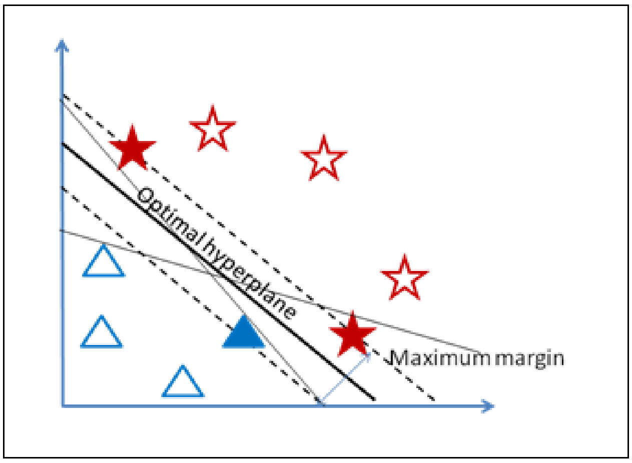
A hyperplane that goes along the training examples will be sensitive to noise while the optimum separating hyperplane has to have maximum distance from the training to have best classifying ability. That is why, the optimality criterion is having the maximum margin which means the distance from the separating hyperplane to the closest members of both sides has to be same and maximum.

The hyperplane which separates the class members and has the maximum margin can be defined as

$wx + b = 0$

where: w is normal to the hyperplane.

Support Vectors are the class members which have the shortest distance to the hyper plane, so the purpose of the support vector algorithms is to find those support vectors which alignes the maximum margin of the hyperplane.



The distance between a point x and a hyperplane (w,b) can be expressed as

$$\frac{|w^T x + b|}{||w||}$$

Despite infinite solutions exist for the best seperating hyperplane, the hyperplane function whose discriminant is selected.

$$|w^T x_i + b| = 1$$

the distance from support vectory the hyperplane is defined as

$$\frac{|w^T x + b|}{||w||} = \frac{1}{||w||}$$

So the margin becomes

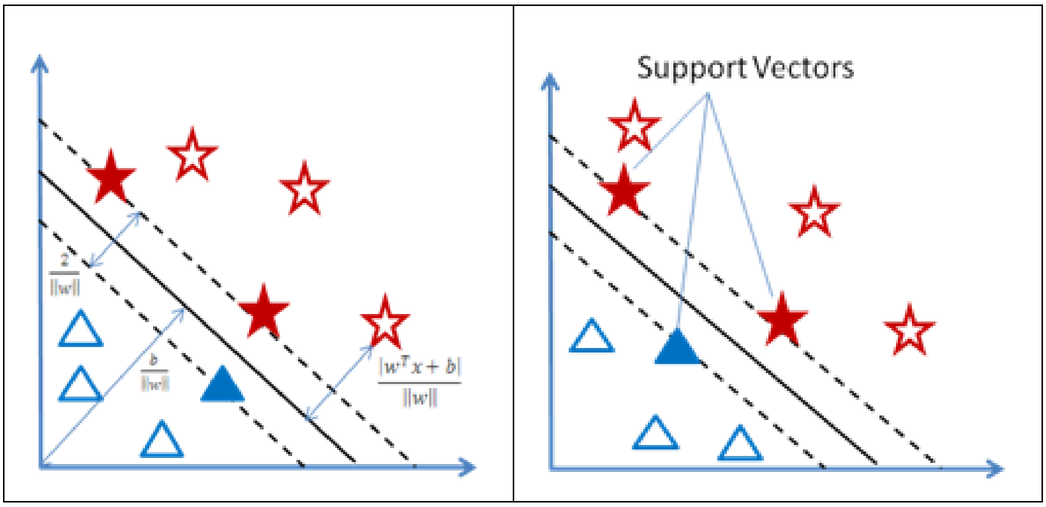
$m=\frac{2}{||w||}$

The optimum margin can be calculated when the following conditions satisfied.

minimize $J(w)=\frac{1}{2}||w||^{2}$

subject to $y_i(w^T x_i + b) \geq 1 \ \forall i$

After solving the optimization problem, those points known as the Support Vectors for which alignes the maximum separating margin on both sides are selected and used for classification purposes as seen figure below.



Multi Class SVM
As SVM algoritm depends on separation of binary classes, how to use it on multi-class problems is still an ongoing issue. There are two main approaches on resolving multi-class problems.[63]

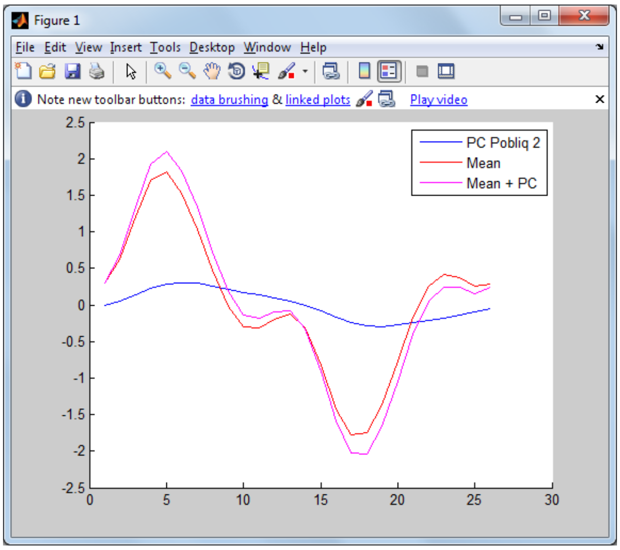
On one vs. all (OVA) classification method, c class classification problem is divided into c binary classifiers. Each of binary classifiers is responsible for separation of a class from all other classes. For example, on each of $g(x)_i$ classifiers, the sample is tested whether it belongs to class c_i. If so, the output is +1, otherwise the output is -1. From all c classifiers, the one which produce +1 result determines the class label. If more than one binary classifier produce +1, tie breaking rules are applied. The most commonly used method on breaking tie is selecting the classifier which has the largest confidence.[64,65]

On one vs. one (OVO) classification method, c multi class problem is divided into $c(c-1)/2$ binary classifiers. Each of binary classifiers is responsible for distinguishing two binary classes. Despite, the number of classifiers is larger than OVA, the total computation cost is almost the same because of lower number of samples. In validation phase, a sample pattern is tested on all classifiers and the class which had the maximum vote determines the class label. When there is tie on class labels, different strategies are used such as weighted voting, classification by pairwise coupling, decision-directed acyclic graph.

APPENDIX E:DISCUSSION OF MEANING OF PRINCIPAL COMPONENTS OF SELECTED FEATURES

The contribution of the PC for selected features were given below. Blue line indicates the PC, red line indicates the mean of the feature.

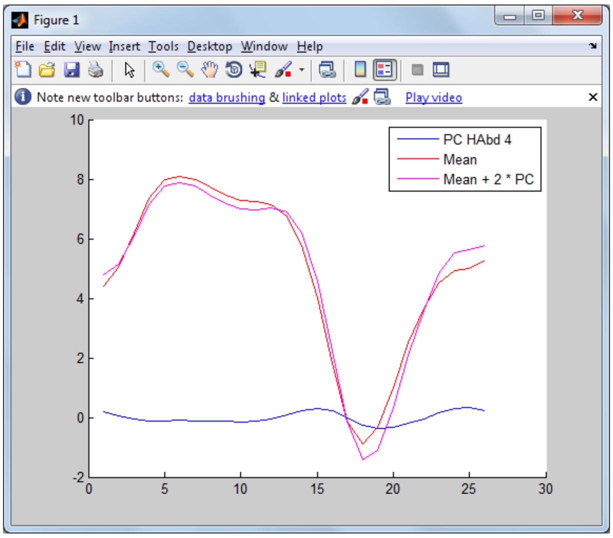
0-1 Data Set



PObliq PC 2;

Increases the amplitude of the waveform for positive coefficient of the PC and shifts right during the stance phase (1-15) for positive coefficient of the PC.

Decreases the amplitude of the waveform for negative coefficient of the PC and shifts right during the swing phase (16-26) for positive coefficient of the PC.

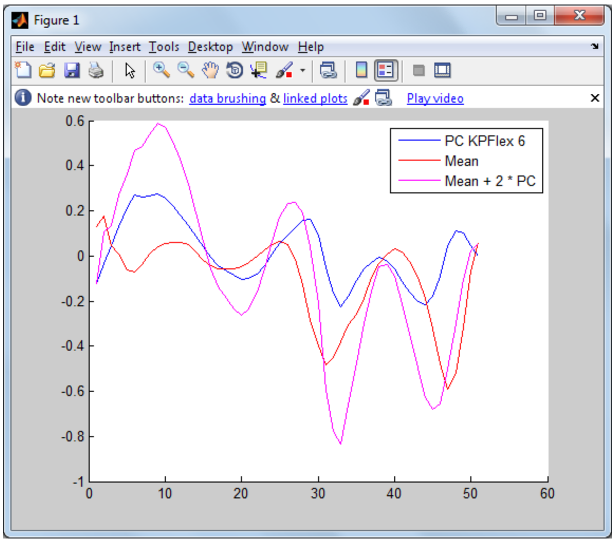


HAbd PC 4;

Decrease the magnitude of the waveform between 1 and 12 for positive coefficient of the PC.

Shifts right the waveform between 13 and 23 for positive coefficient of the PC.

Increase the magnitude after 24 for positive coefficient of the PC.



KPFlex PC 6;

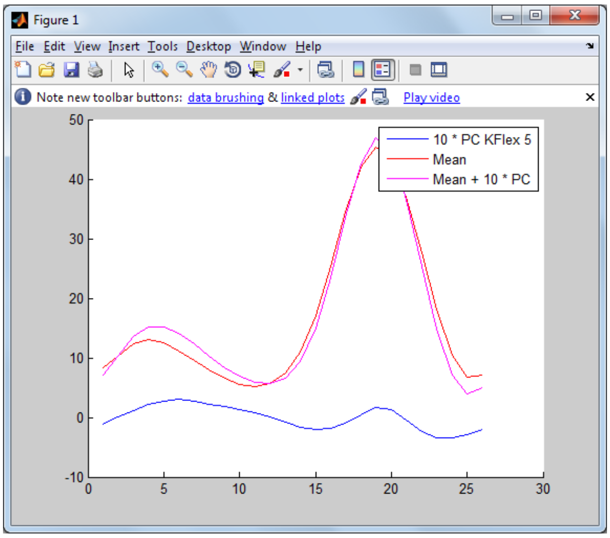
Increase the magnitude of the waveform and shifts right between 1 and 13 for negative coefficient of the PC.

Shifts right the waveform.

Decreases the magnitude between 14 and 38 for negative coefficient of the PC.

Shifts left the waveform between 39 and 51 for negative coefficient of the PC.

0-2 Data Set

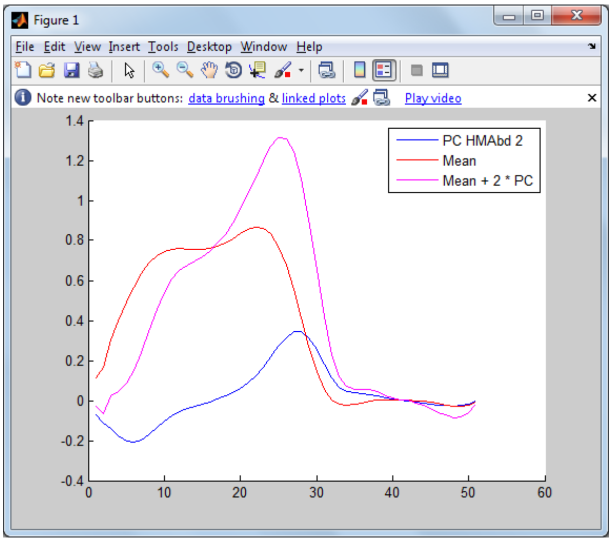


KFlex PC 5;

Shifts right the waveform and increase the magnitude between 2 and 18 51 for negative coefficient of the PC.

Increase the magnitude between 19 and 22 51 for positive coefficient of the PC.

Shifts left the waveform and decrease the magnitude between 23 and 26 for negative coefficient of the PC.

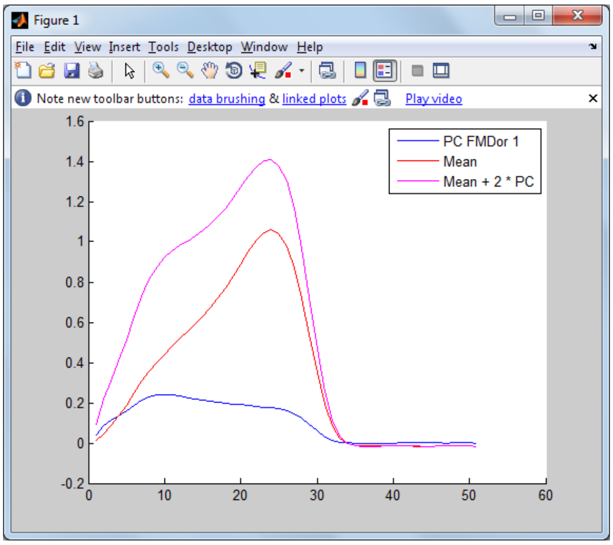


HMAbd PC2;

Shifts the waveform between 1 and 15 for negative coefficient of the PC.

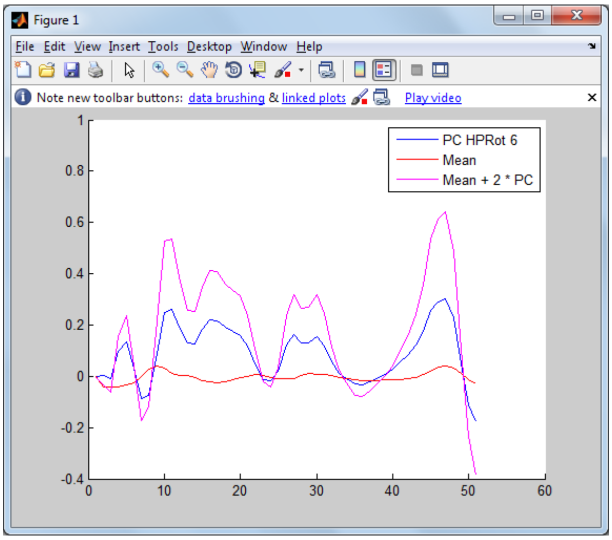
Increase the magnitude and shifts right the waveform between 16 and 40 for positive coefficient of the PC.

0-3 Data Set



FMDor PC 1 increases the magnitude of the waveform during the stand phase for positive coefficient of the PC.

1-2 Data Set



HPRot PC 6;

Increase the magnitude of waveform between 1 and 8 for positive coefficient of the PC.

Decrease the magnitude of waveform between 9 and 10 for negative coefficient of the PC.

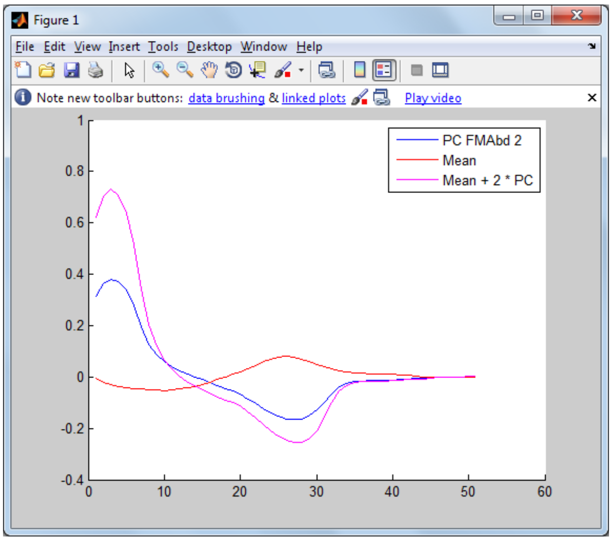
Increase the magnitude of waveform between 11 and 22 for positive coefficient of the PC.

Decrease the magnitude of waveform between 23 and 25 for negative coefficient of the PC.

Increase the magnitude of waveform between 26 and 34 for positive coefficient of the PC.

Decrease the magnitude of waveform between 35 and 38 for negative coefficient of the PC.

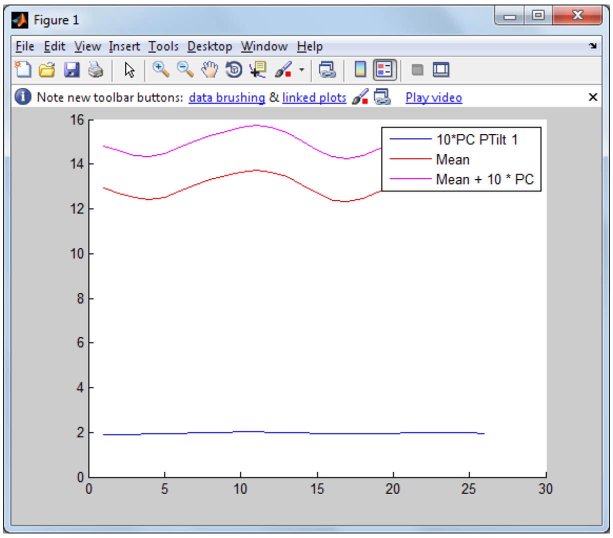
Increase the magnitude of waveform between 39 and 50 for positive coefficient of the PC.



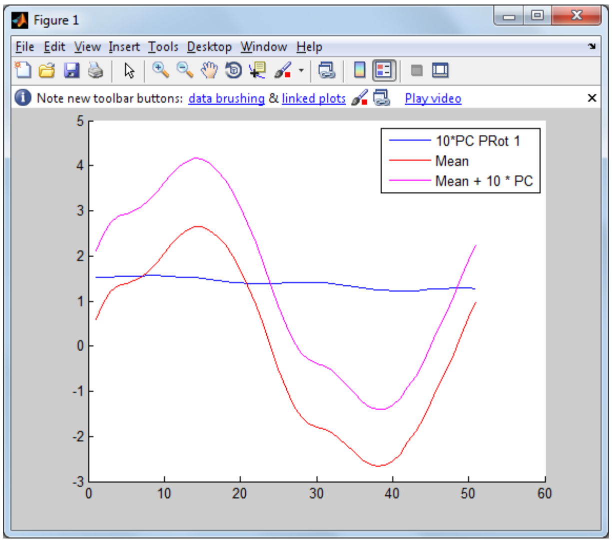
FMAbd PC 2;

Increase the magnitude of the waveform between 1 and 13 for positive coefficient of the PC.

Decrease the magnitude of the waveform between 14 and 51 for negative coefficient of the PC.

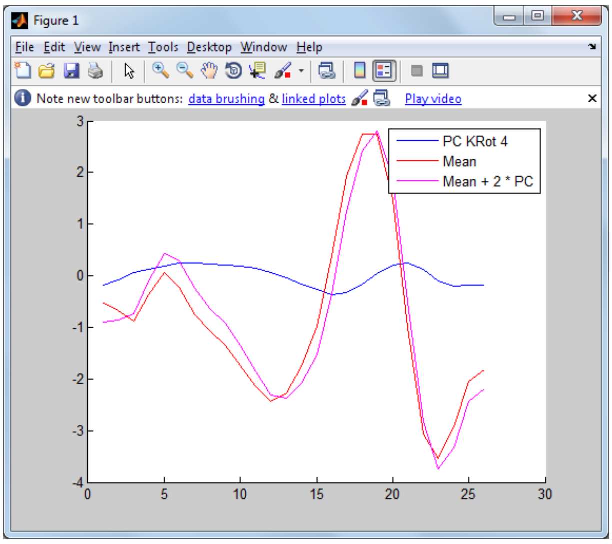


PTilt PC 1 increases the magnitude of the waveform for positive coefficient of the PC.



PRot PC 1 increases the magnitude of the waveform for positive coefficient of the PC.

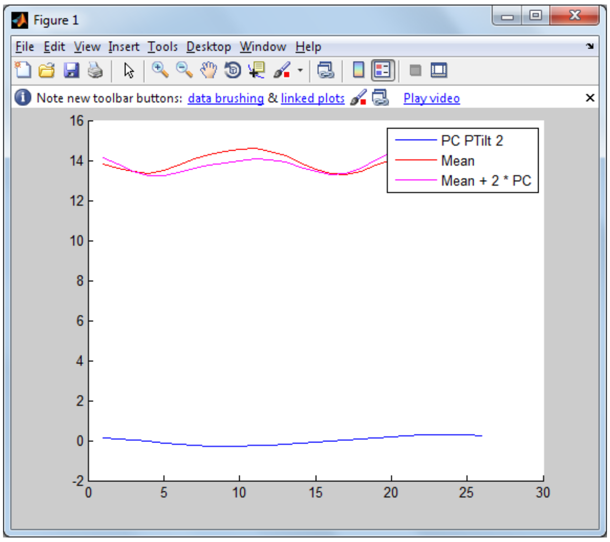
2-3 Data Set



KRot PC 4;

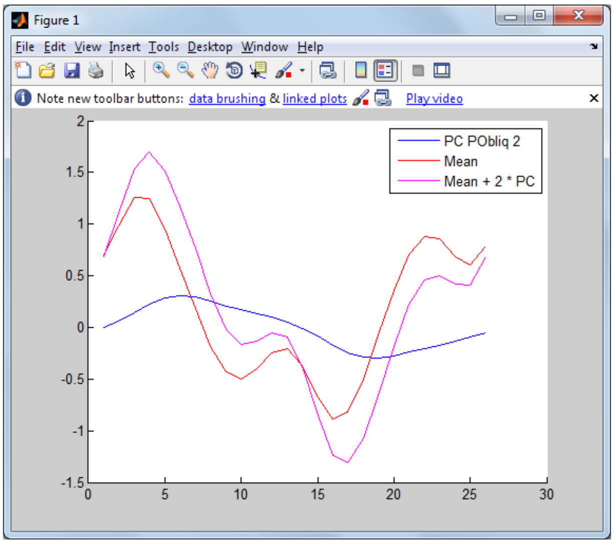
Increase the magnitude of the waveform between 2 and 12 for negative coefficient of the PC.

Shifts right the waveform between 13 and 24 for positive coefficient of the PC.



PTilt PC 2;

Decrease the magnitude of waveform for negative coefficient of the PC.

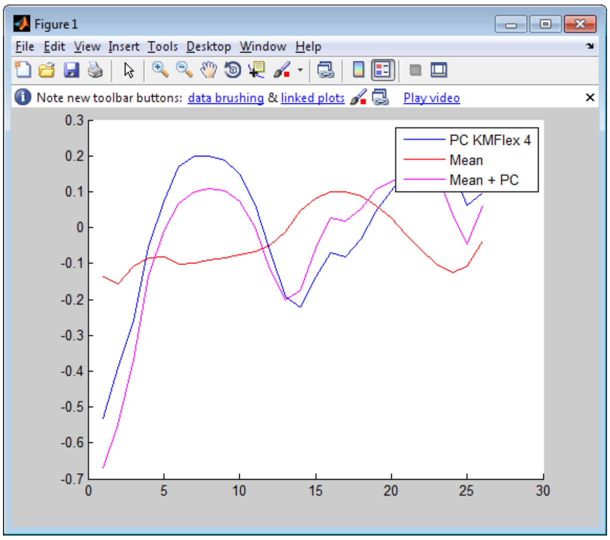


PObliq PC 2;

Increase the magnitude and shift right the waveform during stance phase for positive coefficient of the PC.

Decrease the magnitude and shift right the waveform during swing phase for negative coefficient of the PC.

1-3

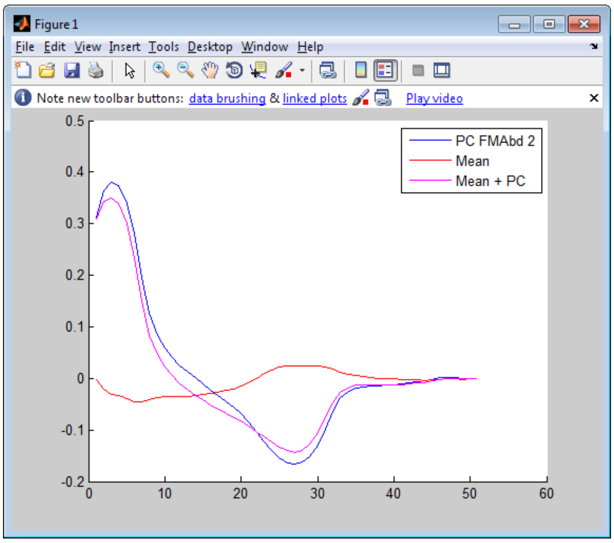


KMFlex PC 4;

Increase the amplitude of the waveform between 1 and 11 for negative coefficient of the PC.

Decrease the amplitude of the waveform between 12 and 21 for negative coefficient of the PC.

Increase the amplitude of the waveform between 22 and 26 for positive coefficient of the PC.



FMAbd PC 2;

Increase the magnitude of the waveform between 1 and 19 for positive coefficient of the PC.

Decrease the magnitude of the waveform between 20 and 51 for negative coefficient of the PC.