UNIVERSITA' DI PADOVA FACOLTA' DI INGEGNERIA

Dipartimento di Ingegneria dell'Informazione

Scuola di Dottorato di Ricerca in Ingegneria dell'Informazione
Indirizzo: Bioingegneria

Ciclo XX

PRINCIPAL COMPONENT ANALYSIS FOR MOTOR SKILLS CHARACTERIZATION AND INDIVIDUAL MONITORING IN SPORTS SCIENCE

Dottorando: Giulia Donà

Supervisore: Ch.mo Prof. Claudio Cobelli

Direttore della Scuola: Ch.mo Prof. Silvano Pupolin

Gennaio 2008

Contents

List of Figures

List of Tables

Sommario

Il presente lavoro è stato svolto nell'ambito della biomeccanica sportiva, scienza che applica le leggi della meccanica e della medicina alla valutazione della performance atletica, per approfondire la conoscenza delle abilità motorie dei soggetti, tramite misure, modellazioni e simulazioni. Tale disciplina si prefigge di soddisfare la crescente richiesta di allenatori, medici ed atleti di ottenere una valutazione quantitativa delle caratteristiche intrinseche alla performance.

La biomeccanica ha svolto un ruolo fondamentale nella comprensione e nell'insegnamento delle tecniche sportive. La modellazione teorica, assieme agli studi sperimentali, ha fornito un valido aiuto nell'approfondire l'analisi dei meccanismi responsabili di determinati fenomeni. L'applicazione della biomeccanica allo sport ha inoltre contribuito a migliorare la performance e ad identificare possibili cause di infortunio.

Negli ultimi anni, si è osservato un notevole incremento nello sviluppo e nell'utilizzo di tecnologie innovative per la ricerca in ambito sportivo. I software attualmente presenti sul mercato offrono inoltre vaste possibilità di visualizzazione e analisi dati. Grazie al sempre più facile accesso a molte strumentazioni di acquisizione dati, ne è divenuto pratico l'utilizzo anche nella valutazione di tecniche sportive. Dai dati ottenuti con le nuove tecnologie, sono però sorte importanti questioni. Innanzitutto, il setup sperimentale deve essere realistico, appropriato allo scopo della ricerca e deve essere mantenuto il più semplice possibile. Si devono inotre adottare tecniche adeguate per il processamento e l'analisi dei dati. Infine, i risultati devono essere riportati in modo semplice, per essere pienamente compresi da atleti e allenatori.

Nella *gait analysis* clinica, si sono già largamente validati numerosi protocolli e l'analisi quantitativa è divenuta un potente strumento nelle decisioni chirurgiche, nel monitoraggio post-operatorio e riabilitativo. In ambito sportivo, il gran numero di discipline e la difficoltà nello standardizzare i movimenti hanno agito da freno ad un uso sistematico di alcune tecnologie potenti, quali, ad esempio, la stereofotogrammetria. Le analisi optoelettroni-

che, sebbene offrano grandi potenzialità, sono infatti ancora ai loro esordi. Devono ancora essere affrontati problemi sostanziali, come il tempo necessario all'acquisizione dati, la gestione degli attrezzi, i costi, la mancanza di modelli specifici per i singoli soggetti, ecc.

Compito principale degli allenatori è l'ideazione e la strutturazione di programmi di allenamento che producano un progressivo miglioramento dell'atleta. Un ruolo importante ricopre anche la prevenzione di infortuni. E' quindi evidente che uno degli obiettivi della ricerca in ambito sportivo debba essere l'identificazione delle caratteristiche peculiari e delle strategie motorie più efficienti per il singolo atleta. Il monitoraggio di più atleti è infatti la base per la descrizione quantitativa accurata del gesto motorio in analisi e la valutazione dei diversi livelli di abilità degli atleti. La conoscenza delle attitudini e delle carenze dei singoli atleti deve poi concorrere a differenziare programmi di allenamento mirati. Inoltre, per ottenere un'affidabile caratterizzazione del singolo soggetto si ha la necessità di svolgere un monitoraggio longitudinale: l'atleta deve essere confrontato con se stesso in periodi diversi del periodo di allenamento.

Nel cercare di cogliere le caratteristiche individuali, la ricerca non deve focalizzare l'attenzione sulla miglior performance dell'atleta. Al contrario, deve essere individuata la "modalità tipica" della performance dell'atleta. Con il termine modalità tipica si intende la ripetizione di un movimento con un alto grado di certezza e con la massima efficienza. Per cui, la comprensione delle caratteristiche motorie individuali risulta strettamente legata alla valutazione di un ampio numero di ripetizioni di un movimento. I dati derivanti da analisi ripetute sono fortemente affetti dalla presenza della biovariabilità. Infatti, atleti diversi svolgeranno un movimento in modi differenti. Per di più, anche atleti di massimo livello non sono in grado di riprodurre un pattern motorio sempre nello stesso identico modo, anche se si allenano da molti anni. Le variabilità inter- ed intra- soggetto giocano quindi un ruolo fondamentale nella valutazione delle abilità motorie e la loro azione sui dati deve essere tenuta in considerazione.

Nella biomeccanica sportiva e nella *motion analysis*, anche protocolli sperimentali semplificati producono spesso un gran numero di dati. E' quindi necessario un notevole sforzo per trovare una struttura nei dati, identificare le *features* più caratteristiche e presumere se un andamento sia rappresentativo dell'atleta o meno. Solitamente i risultati vengono interpretati soggetivamente, a partire da un vasto numero di variabili altamente correlate e tempovarianti. Le tecniche standard di analisi dati (media, deviazione standard, ecc.) falliscono nell'estrarre l'informazione significativa da un insieme numeroso di variabili cinematiche e dinamiche. Al contrario, l'analisi statistica multivariata si è dimostrata un potente mezzo per eliminare la collinearità dai dati e facilitarne la comprensione, evidenziando solamente le strutture essenziali nascoste in essi. Tra le tecniche statistiche multivariate, le trasformazioni lineari sono computazionalmente più semplici da gestire. Tra di esse, l'analisi delle componenti principali (PCA) è risultata essere molto efficace nello studio del movimento umano. E' stata infatti utilizzata in letteratura per identificare gruppi di variabili correlate, discriminare differenti pattern di cammino o valutare i cambiamenti dovuti a patologia, infortunio od intervento. Ciò nonostante, la PCA non è stata ancora sfruttata nell'ambito della biomeccanica sportiva, per le valutazioni di stato dell'atleta, prevenzione degli infortuni, allenamento, ecc.

Scopo di questa tesi è valutare le potenzialità dell'uso della PCA nel ridurre e interpretare dati derivanti da un'analisi sportiva, tenendo in considerazione la variabilità insita in essi. Si è scelto di utilizzare la marcia come strumento di analisi, perchè è un gesto motorio che presenta proprietà biomeccaniche e coordinative altamente tecniche e ripetibili. Sono stati utilizzati un sistema optoelettronico ed una pedana di forza per acquisire e stimare la cinematica e la dinamica di sette marciatori di livello internazionale. Sono stati processati i dati derivanti da più ripetizioni del gesto motorio.

L'analisi delle componenti principali è una tecnica statistica che riduce la dimensione dei dati originali, mantenendo l'informazione più importante in essi contenuta. Le variabili vengono rappresentate in un numero ridotto di componenti, che tengono conto di gran parte della variabilità originale. In questa tesi sono state analizzate tre possibili applicazioni di analisi delle componenti principali: tradizionale (*t-PCA*), funzionale (*f-PCA*) e a due fasi (*2-PCA*). Un ulteriore obiettivo è stato quello di valutare i vantaggi e gli svantaggi delle tre tecniche nel rispondere a differenti quesiti, dal momento che queste tecniche non sono ancora state ampiamente adottate in ambito sportivo.

Innanzitutto, si è ottenuta una caratterizzazione generale della biomeccanica della marcia, in modo da avere un quadro completo del movimento in analisi. Si è cercato poi di ottenere una caratterizzazione robusta e completa della strategia motoria di ogni atleta. Tutte le tecniche utilizzate hanno permesso di individuare relazioni tra le variabili in analisi e la performance di marcia. E' stato possibile identificare i principali elementi che distinguono gli atleti in base al loro livello di abilità. Si sono inoltre descritte le peculiarità tecniche e coordinative di ognuno di essi. Infine, si è riportato un esempio di monitoraggio longitudinale. La PCA è stata infatti applicata ai dati derivanti da due sessioni di test successive, in modo da valutare i miglioramenti ottenuti grazie all'allenamento.

Questo studio ha cercato di mostrare come l'analisi delle componenti principali possa rappresentare un valido strumento per caratterizzare le abilità motorie degli atleti ed effettuare un monitoraggio longitudinale. La PCA si è rivelata utile anche nell'ottenere importanti informazioni su comportamenti motori che potrebbero essere causa di infortuni. Si è adoperata un'attenzione particolare nel cercare una connessione tra un approccio matematico complesso e teorico quale la PCA e la sua applicazione pratica. Si è cercato infatti di dare un'interpretazione biomeccanica dei risultati ottenuti, al fine di rendere l'informazione comprensibile anche agli atleti.

Summary

Sports science defines the field in which this research was carried out. Sports biomechanics is the science that applies the laws of mechanics and physics to athlete performance, in order to gain a greater understanding of motor skills through measurement, modeling and simulation. It operates for meeting the growing demands of coaches, physicians and athletes, to quantitatively assess the essential characteristics of performance. Biomechanics has played an important role in the understanding and coaching of sports techniques. Theoretical modeling and experimental studies have helped in going more deeply into the mechanisms responsible for an observed phenomenon. The application of biomechanics in sports has also contributed to improvements in performance and to the understanding of possible causes of injury.

Throughout the last years, increasing efforts in developing and using innovative technologies in sport science has been observed. Moreover, currently available software offer numerous options for the visualization and analysis of biomechanical data. As methods for data collection have become more widely available, it has become practical to use them in the evaluation of sports technique. Some important issues have been raised by these investigations. First, the experimental setup should be realistic, appropriate for the purposes of the research, and should be kept simple. Then suitable approaches should be adopted in data processing and results analysis. Finally, results should be reported in an easy way, to be fully understandable by athletes and trainers.

In clinical "gait analysis", standard protocols have been widely validated and quantitative analysis has become a powerful tool for surgical decision and for post-operative and rehabilitative monitoring. In sports field the great amount of disciplines and the difficulty in standardizing movements have acted as a brake on the systematic use of powerful techniques like stereophotogrammetry. These methods, although offering great potentialities, are still at their beginning. Substantial issues are still to be investigated, such as time needed for data collection and data analysis, handling of the equip-

ment, costs, lack of subject-specific models, etc.

The main task for coaches is to identify and carry out training programs that determine progressive improvement of athletes. Furthermore, all injuries should be prevented, too. Therefore, one of the purposes of sports research should be the identification of the peculiar characteristics and of the most proficient strategy for each athlete. The monitoring of a group of athletes should be the base for an accurate quantitative assessment of the movement under analysis and for the identification of different skill levels. Then, the knowledge of single athlete's abilities or deficiencies should help coaches to adjust individual training programs. Moreover, a reliable characterization of the subject should pass through longitudinal monitoring: the athlete might be compared with himself in different times during the training season.

When trying to capture the individual characteristics, the research should not be focused only on the best performance of an athlete. In contrast, studies should analyze the individual "typical mode" of the athletic performance. With the term typical mode it is intended the repetition of a movement with a high degree of certainty and maximum proficiency. Hence, the comprehension of individual motor behaviors is strictly related to the evaluation of a large number of repetitions of a movement. Results of this kind of analysis are strongly affected by the presence of motor biovariability. Different athletes perform the same task in different ways. Moreover, even elite athletes can not reproduce identical movement patterns, after many years of training. Thus, inter- and intra-individual variability in movement patterns play a fundamental role in sports skills and its influence on data should be accounted for.

In sports biomechanics and in motion analysis, even simplified protocol setups often provide a large amount of results. Remarkable efforts are required in finding a structure in the data, discovering the most characteristic features and predicting whether a pattern is representative for the athlete's skill description or not. Results are commonly interpreted subjectively from a large number of highly correlated, time-varying and constant variables. Standard data analysis techniques (mean, standard deviation, etc.) lack in extracting significant information from a large amount of kinematic and kinetic data. In contrast, multivariate statistical analysis has proved to be a powerful tool to eliminate collinearity and to facilitate the analysis, presenting only the essential structures hidden in the data. Among multivariate statistical techniques, linear transformations are computationally easier to handle. Among linear transformations, principal component analysis (PCA) has showed to be extremely effective in the study of human motion. It has been usually adopted to identify groups of inter-related variables, different walking patterns or gait forms, or to evaluate changes due to pathology, recovery or intervention. Nevertheless, PCA has not been used in sports biomechanics, to assess athlete's training status and recovery needs, tracking injuries, managing training programs, etc.

The purpose of this thesis was to investigate the use of principal component analysis for reducing and interpreting sports motion data, while accounting for their original variability. Race walking was chosen as the mean of investigation, because it is a motor task that presents peculiar biomechanical and coordinative demands. An optoelectronic system and a force plate were used to collect and estimate kinematics and kinetics of seven race walkers of international level. Several race walking repetitions were acquired and kinematics and kinetics variables were processed.

Principal component analysis summarized the most important information in the data, by representing the variables in a limited number of components that explained most of data variance. Data underwent three different applications of PCA: traditional (*t-PCA*), functional (*f-PCA*) and two-stage (*2-PCA*).

A further objective of this thesis was to evaluate the advantages and disadvantages of the three methods in solving different challenges, because these techniques have not been widely adopted in sports research. A general characterization of race walking biomechanics was pursued, in order to get a full comprehension of the movement under analysis. Then, a robust and complete characterization of the single athlete's performance strategy was given. All the three methods allowed the exploitation of the relationships among multiple measures in the analysis of race walking data. The most important factors that distinguish athletes according to their skill levels were found out. Moreover, the peculiar technical and coordinative characteristics of each athlete were widely described. Finally, an example of longitudinal monitoring was described. Motion analysis, combined with PCA, was used on data from two subsequent testing sessions, to identify the main improvements caused by training.

This study tried to show how principal component analysis could represent a valuable tool for motor skills characterization and individual monitoring. It gave also important information about motion behaviors that might be primarily responsible for injury. Moreover, a special effort was spent in finding a connection between a complex and theoretical mathematical approach (PCA) and its practical application. The biomechanical interpretations of the statistical results was intended to make information intelligible for practitioners.

Chapter 1

Introduction

1.1 Sports biomechanics

The science and practice of sport are becoming objects of increased interest. Physicians, coaches, athletes and even recreational sportsmen are showing rising curiosity in the quantitative assessment of the athlete's motor characteristics. Qualitative analyses of human movement are commonly adopted both for evaluating athletes' abilities and testing training effectiveness. They have often been based on a subjective interpretation of movement. Qualitative analyses mainly consist in observation, evaluation and intervention. Competition results, field tests and visual inspection have traditionally been adopted to characterize the athlete's abilities in performing a motor task. These methods, although immediate and very user-friendly, can not provide an exhaustive description of how the performance is carried out. Moreover the perception and the correctness of the movement are strictly connected with the experience and the coaches' knowledge. The need of trainers for increasing athletes' performance and preventing injuries is the basis for research development in improving the quantitative assessment of motion.

During the last decades, researchers have driven sports biomechanics from its descriptive phase to a more analytical one [80, 106]. The laws of mechanics and physics were applied to human performance, in order to gain a greater understanding of performance in athletic events through modeling, simulation and measurement. Biomechanical studies primarily cover issues of performance enhancement, injury prevention and safety regarding elite, leisure and rehabilitative activity. They allow coaches to rationalize the instructions they give and to provide the basis for improving performance [69].

Three-dimensional motion analysis represents a very useful tool to understand how sports skills are performed. Powerful and reliable devices for motion analysis are currently available [3, 14, 22, 36], together with many different multi-segment models of the human body and software tools for simulating body motion.

Despite this great improvement in theoretical knowledge and the availability of instrumentation for sports biomechanics, there is a lack in the definition of appropriate experimental setup, data elaboration and data analysis. In clinics, quantitative analysis has become a powerful tool for surgical decision and for post-operative and rehabilitative monitoring. In particular, in gait analysis, the interaction among clinicians, therapists and engineers has brought to the definition of standard protocols and the sharing of a "common language". In sports field the great number of disciplines, the complexity of athletic movements and the heterogeneity of survey questions [7, 91] prevented such a standardization. The scientific demands of validity, reliability and accuracy have to front the more practical issues like complexity, range of motion to be analyzed, time needed for data collection and data analysis, handling of the equipment, costs and invasiveness. Moreover careful information management is required to optimize the process of data elaboration and feedback of biomechanical information to coaches and athletes [5, 7, 44, 72, 91, 97].

Although more controlled experimental studies and suitable statistical data elaborations have been required from some authors [7, 91, 125], they remain a rarity. Some procedural issues in studies comparing different athletes' techniques have not still been resolved. A need for comprehensive, common guidelines on experimental protocols and data proceeding and analysis remains.

1.1.1 Issues in Sports Biomechanics

The aim of sports research should be the identification of the peculiar characteristics and of the most proficient strategy for each athlete. Since every athlete is characterized by his own abilities and deficiencies, trainers and coaches might get more effective results by applying and monitoring individual training programs rather than using the same strategy for every athlete.

Variability

Significant differences in individual performances have been always observed in experimental studies in which different athletes were compared. Such fluctuations among different individuals are commonly called "inter-subject" variability. This variability hardly lend support to the idea of an optimal performance model, or technique. Movement variability is present both in sports in which the task is predominantly speed dependent, or requires speed and accuracy, and in repetitive cyclic sports. Hence, competition results or experimental outcomes are questionable unless the reliability and objectivity of the data are assessed and reported.

Focusing attention on individual analysis, also repeated trials of the same subject present results included within a range of deviation. In fact it is well known that even elite athletes can not reproduce identical movement patterns after many years of training [7, 34, 47, 48, 49, 90, 91]. Intra-individual variability in movement patterns plays a fundamental role in sports skills and particular attention has to be paid to it.

Movement variability is not exclusively due to neuromuscular system or measurement "noise", but it is functional [8, 49]. The functional importance of variability has been recognized mostly by experts in the ecological theory. This theory holds that all movements and actions are influenced or constrained by the environment. Environmental information is necessary to shape or modify the characteristics of movement to achieve specific actions or tasks. Thus variability is seen as allowing environmental adaptations, reducing injury risk and facilitating changes in coordination patterns [7, 8]. Variability in movement is particularly important in many sport skills for which the adaptability of complex motor patterns is necessary [49].

Moreover, group designs concern the need for averaging the results. Since different athletes perform the same task in different ways, there is no optimal average movement pattern for athletes as a whole [7, 8, 91]. Taking an average may obscure important individual differences. Therefore more emphasis should be placed on individual designs, to address issues such as individual "signature" of the movement or the optimization of performance [8, 91].

Many problems arise when one tries to capture individual characteristics. First, the research should not be focused only on the best performance of an athlete. In contrast, research should analyze the individual "motor skill", which can be defined as the ability to obtain the goal with a high degree of certainty and maximum proficiency [105]. Research results should be representative of the "typical mode" of the athlete's performance [77]. One definition of a representative mode is "the central tendency score" [66]. This definition requires that variability, i.e. "the degree of difference between each individual score and the central tendency score" [111], is measured. To obtain these measures, several trials need to be recorded [77, 90, 91]. Trial size is influenced by many factors: experimental errors, robustness of the assumptions of the analysis method, research design, consistency of athletes' responses, etc. [77, 91].

Data Analysis

When many trials of the same movement are performed, measurements do not generate a single curve, but a family of curves, each one slightly different from the other. There is the need to find a structure in the data, discovering the most characteristic features and predicting whether a pattern is representative for the athlete's skill description or not. Hence, two main needs have to be satisfied: data reduction and interpretation. The first demand has the intent to eliminate collinearity and to simplify data. The second one concerns the need to obtain a conceptually meaningful summary of the data. Approaches for the modeling and recognition of motion data may be divided [18] into data fitting, feature location and multivariate statistical analysis techniques. Focusing on the multidimensionality of the data set, one of the methodology most frequently applied to human movement patterns is multivariate analysis technique [26]. Multivariate data analysis is a fundamental way to find suitable representations of measured data, facilitating the analysis and presenting only the essential structures hidden in the data [64].

Among multivariate statistical techniques, linear transformations are computationally easier to handle [1]. Linearity reduces the complexity of transformations and of computational resources. Moreover, it facilitates the interpretation of the results. Due to these advantages, linear transformations have been used in many applications, and various techniques for finding linear transformations have been developed. Principal component analysis, factor analysis, and projection pursuit are examples of well-known techniques for finding linear transformations.

Principal component analysis (PCA) is a technique that has been shown to be extremely effective in the study of human motion. First, PCA explicitly deals with variance in the data. Given the relevance of variability in sports measurements, it might be a suitable tool to summarize and compare data in a way that reflects the true nature of their variability [18].

Works asserting the usefulness of PCA for dimensionality reduction and subsequent interpretation can be widely found in literature. Moreover, this statistical technique demonstrated a great ability in analyzing entire waveforms and thereby retaining valuable temporal information. Since PCA revealed to be useful for clinical applications, its potentialities were investigated also for a sportive discipline.

1.2 Aims

The purpose of this thesis was to investigate the use of principal component analysis (PCA) for reducing and interpreting race walking data, retaining most of their original variability. In fact, PCA summarizes the most important information in the data, by representing the variables in a limited number of components that explain a maximal amount of variance.

Race walking was chosen as the mean of investigation, because it is a highly technical and very reproducible motor task. Data underwent three different applications of PCA: traditional (*t-PCA*), functional (*f-PCA*) and two-stage (*2-PCA*).

Since principal component analysis has not been widely adopted in sports research, an objective of this thesis was to extensively explore and assess the application of the PCA to athletic movements. A second objective was to evaluate the advantages and disadvantages of the three methods in solving different challenges for athletes' skill characterization and longitudinal monitoring.

The final purpose was to evaluate how a complex and theoretical mathematical approach (PCA) can find practical application by matching results with field needs (motor skills characterization) and by making information intelligible for practitioners.

1.3 Structure of the thesis

A brief description of peculiar characteristics of race walking is discussed in chapter two. Basic and essential information about the three types of PCA applied to data will be addressed in the third chapter. The information about the recruited subjects, the experimental setup, the applied protocol and data processing are in chapter four. Results regarding athletes' skill characterization, coordinative description and an example of longitudinal monitoring are presented in chapter five. A general discussion will follow at the end.

Chapter 2

Race walking

This chapter explains why race walking was chosen as the mean of investigation. A brief description of the discipline is included to better understand its biomechanical characteristics.

2.1 Description

An official definition of race walking was established to formally differentiate it from running. This definition evolved over several decades as officials attempted to better define the sport. Race walk [115] judging requires the observation of walkers competing at fast speeds. In 1995 as Rule 230.1 of the International Association of Athletics Federations (the IAAF) [58] was written:

> *Race Walking is a progression of steps so taken that the walker makes contact with the ground so that no visible (to the human eye) loss of contact occurs. The advancing leg shall be straightened (i.e., not bent at the knee) from the moment of first contact with the ground until the leg is in the vertical upright position.*

In other words, two rules must be strictly respected to perform a correct action: first the race walker must maintain continuous contact with the ground during the progression of steps; second, the knee of the walker's advancing leg must remain straight, without bending, from the moment of the first contact with the ground until the leg is in the vertical position. An oversight in the observation of these rules during the competition implies sanctions or disqualification [58].

The race walking action can be divided into four principal phases:

- the front leg support phase - FSP,
- the rear leg support phase - RSP,
- the double support phase - DSP,
- the swing phase - SP.

The stance phase can be further subdivided in six main stages, as reported in Figure 2.1.

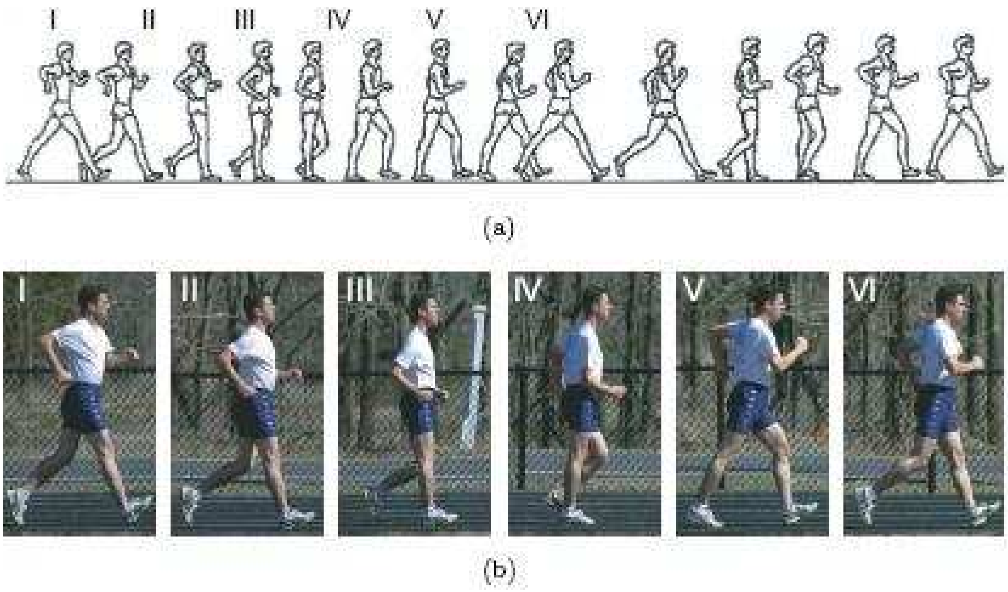


Figure 2.1: Race walking technique during the complete gait cycle (a) and main stages of stance phase (b) [90].

The FSP goes from phase I to III: it begins at heel strike and ends when the supporting leg passes beyond the vertical projection of the center of mass. There are some fundamental differences between normal walking and race walking. The functional lengthening of the lower limbs at the moment of heel strike occurs performing: more plantar-flexion of the ankle of the rear limb; more extension of the knee and dorsi-flexion of the ankle of the front limb [78]. At the beginning the body is in the double-support phase; the body weight is distributed on the heel of the front foot and the big toe of the back foot. By landing on the heel, with the toe up, the race walker with enough forward momentum can ride the body over the locked leg [115]. This phase represents a rapid transition from load absorption to propulsion, therefore the athlete should minimize the loss of progression velocities by performing a "smooth" action. The rear leg then thrusts forward after the toe of the back foot pushes off the ground. Afterwords the body moves forward, over the front leg. At this point walkers tend to violate the rules of race walking: the leg must remain straightened until it is in the vertical position [81, 115].

The RSP begins when the stance leg passes the vertical upright position and ends with toe-off (IV-VI). In this phase the knee may be bent. However, if the leg is straightened longer, an extra thrust forward is get by pushing off the rear foot. Ideally, the leg should remain straightened until the heel of the rear foot lifts off the ground. After push-off the leg swings forward to complete the gait cycle [91, 115]. This passage should be performed by driving the knee as low to the ground as possible, in order to minimize energy expenditure, to preserve the horizontal speed and to give a better appearance of correct technique [90, 91].

As the front leg leaves the ground, the rear leg is already in contact with the ground (I and VI). This transition between RSP and FSP is the double support phase (DSP), which is necessarily present, to respect the first of the two main rules of race walking. The legs do not create a symmetrical triangle in the sagittal plane: the rear foot is more behind the body than the front foot is ahead of it. This derives from the two chances of generating propulsion which proceed concurrently: as front leg creates momentum by driving forward, the rear leg pushes back against the ground, launching the body forward by way of a powerful vaulting effect. Keeping the rear foot on the ground as long as possible by rolling up onto the toe at push-off will maximize this leverage. This action is also achieved through proper hip and pelvis action: the increased obliquity of the pelvis in the frontal plane ("hip drop") can be used to further extend the driving leg and delay toe off.

The last part of the race walking cycle consists of the swing phase, when the foot has no contact with the ground, swings and then prepares to approach back the ground. The hip is flexed in midswing, whereas it is extended throughout the rest of swing [78]. Just before contact the leg is swinging forward, straightening.

Hip and pelvic motions are fundamental in race walking (Fig.2.2). The hips are the body's primary source of forward locomotion. Hips drop and roll while twisting back and forth. Actively swinging the hip forward increases stride length behind the body. Thus a strong front-to-back hip action facilitates both the propulsive phases of the stride. This allows the legs to move faster and easier and gives the race walker a longer stride.

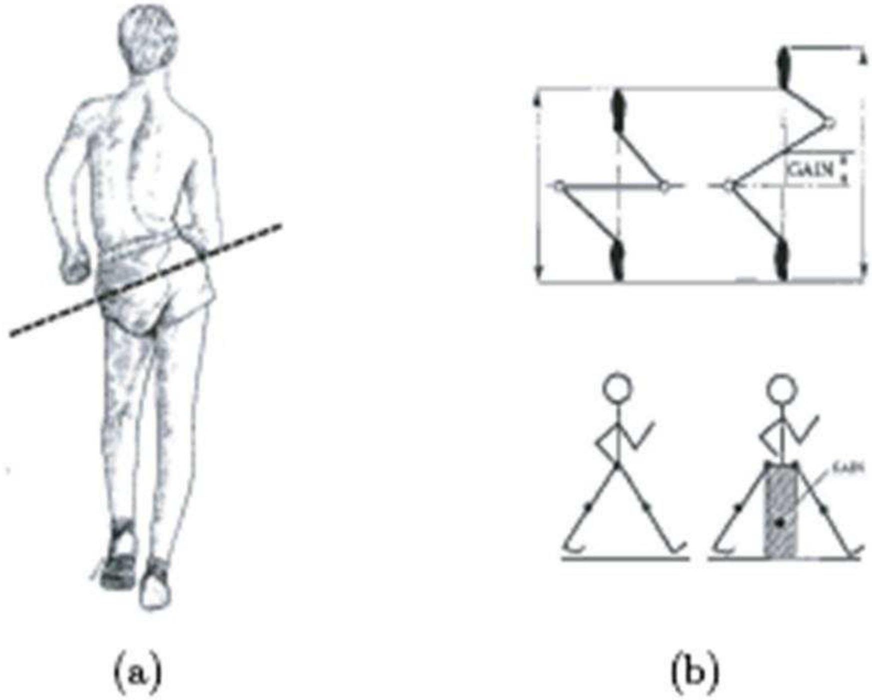


Figure 2.2: Race walking technique concerning pelvic rotations in the frontal (a) [78], horizontal and sagittal (b) planes [90].

The movement of hips has a strong influence on the position of the center of mass (COM), a fundamental aspect in race walking technique. While in normal walking, the center of mass (COM) oscillates, in race walking the pronounced arm swing and the emphasis on strong toe off replace much of the oscillation and level the path of COM as it moves forward [13, 90, 91, 98, 115]. Moreover the straightening of the supporting leg and the compensatory lowering of the opposite hip and ipsilateral shoulder facilitate in maintaining a linear progression of the center of mass. There is less rising and falling, and higher race walking velocities are possible. The locked position of the knee is compensated by an increase in the obliquity of the pelvis in the frontal plane (Fig.2.2(a)). Furthermore, greater pelvic rotations in the horizontal plane (Fig.2.2(b)) concur in improving the step length. The hip movement is crucial also in maintaining a correct foot placement: in race walking feet should line up one behind the other. When trying to walk this near straight-line placement without using the hips, an unneeded stress is placed across the knee. Moreover a correct hip action allows to increase the stride length and to make it as smooth as possible.

2.2 Motivation

Race walking was chosen as the mean of investigation because of its specific constraints. The restrictions imposed by the main rules result in very particular biomechanical and coordination demands. In fact, the rules restrictions add further control over the execution and make race walking appear as rather stereotyped. Therefore, the choice of a very repeatable movement seemed a good basis for gaining more insight into PCA potentials in the description of a sportive action. In fact kinematic and kinetic variables of race walkers of high level are very similar and can not be distinguished through standard data analysis techniques [90].

Moreover, race walking has strong analogies with normal walking, one of the most documented movements in literature. Principal component analysis, together with some other complex statistical techniques, has already been adopted on gait data [18, 19, 20]. This provides an excellent methodological support for knowing how most suitably applying the technique and processing data.

Furthermore, a review of the literature revealed that little has been documented about biomechanics of race walking [13, 90, 91, 78, 81]. This discipline is not an inborn motor strategy, because at the progression speed that race walkers are able to achieve, the man would naturally turn from walking to running [13, 78]. The chance of a straight comparison with normal walking characteristics helps in deepening the knowledge of race walking mechanical and physiological peculiarities. In fact, it gives preliminary indications on the deviations from normal patterns, injury prevention, rehabilitation monitoring, etc.

2.3 A glossary of biomechanical terms for race walking

The following biomechanical terms will be used to describe race walking biomechanics in this thesis [115]:

- *Toe Off* → when the rear foot loses contact with the ground;

- *Heel Strike* → when the foot makes contact on the heel;
- *Step* → from toe off to heel strike of the same foot;
- *Race walking Cycle* → the interval of time from toe off of one foot to toe off by the same foot. It has two phases, swing and stance;
- *Stance Phase* → when body weight is supported by the legs;
- *Single Support* → when one foot is in contact with the ground;
- *Double Support* → when two feet are in contact with the ground;
- *Swing Phase* → when the leg is moving forward during a step;
- *Extension* → body part going from a bent to a straight position; where two or more bones comprising a joint move away from each other;
- *Hyperextension* → the extension of a part of the body beyond normal limits;
- *Flexion* → body part going from a straight to a bent position; where two or more bones comprising a joint move closer to each other;
- *Vertical Position* → when the leg is directly under the center of mass.

Chapter 3

Principal Component Analysis

This chapter provides some details of principal component analysis. Special attention is reserved to the description of peculiarities of the three methods adopted to analyze data: traditional (*t-PCA*), functional(*f-PCA*) and two-stage (*2-PCA*).

3.1 Mathematical and geometrical explanation

Principal component analysis (PCA) is a technique that has been shown to be extremely effective in the study of human motion. PCA is applied to extract the relevant information from high-dimensional data sets by considering only those principal components that sufficiently explain high fractions of the entire data set in terms of its spread or variance. PCA is an old tool in multivariate data analysis, widely adopted in movement analysis [56, 87]. It has been substantially used to identify groups of inter-related variables, to recognize different walking patterns or gait forms, or to evaluate changes due to pathology, recovery or surgery.

More in detail PCA is an approach to find the most informative or explanatory features in data, without needing a priori knowledge. It accomplishes this by computing a new, much smaller set of uncorrelated variables (Principal Components - *PCs*), which best represents the original data-set. Each new variable is a linear combination of the original ones.

The first principal component (*PC1*) is the linear combination of the original variables which accounts for the maximum amount of variance in a single direction. It is the line of best fit through the data, and the residual variance about this line is then a minimum for the data set. The second principal component (*PC2*) is that line which is orthogonal to the first one and accounts for the maximum amount of the remaining variance in the data. All the principal components are orthogonal to each other, so there is no redundant information. The first two components therefore represent the plane of best fit through the data. All remaining principal components are defined similarly, so that the lowest order components normally account for very little variance and can usually be ignored.

Geometrically, principal component analysis can be considered as a rotation of the axes of the original variable coordinate system to new orthogonal axes, called "principal axes", such that the new axes coincide with the directions of maximum variation of the original observations. An example is reported in Fig.3.1.

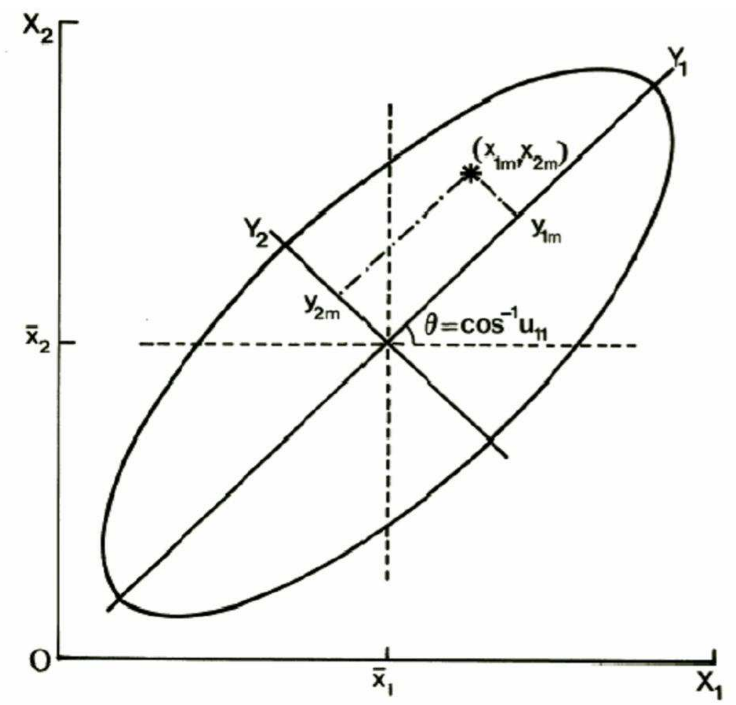


Figure 3.1: Example of the geometrical description of principal components. Y_1 and Y_2 are the principal axes; the points y_{1m} and y_{2m} give the principal component scores for the observation $x_1 = (x_{1m}, x_{2m})$; the cosine of the angle θ between Y_1 and X_1 gives the first component of the eigenvector corresponding to Y_1.

The point y_{1m} is the projection of the point $x = (x_{1m}, x_{2m})$ onto the axis defined by the direction Y_1. This axis has the property that the variance of the projected points $y_{1m}, m = 1, 2, ...n$ is greater than the variance of the points when projected onto any other line or axis passing through $(\overline{x_1}, \overline{x_2})$. The point y_{2m} is the projection of the point $x = (x_{1m}, x_{2m})$ onto the axis defined by the direction Y_2, orthogonal to Y_1. y_{1m} and y_{2m} are named principal component "scores".

Mathematically the principal component analysis is conducted in a sequence of steps, with somewhat subjective decisions being made at many of these steps.

Singular Value Decomposition

Given a data set represented as a matrix X of n rows and m columns

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{3.1}$$

the sample mean is computed for each column j:

$$\mu_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}, \quad j = 1, 2, ...m \tag{3.2}$$

Therefore, X can be centered to form X^* :

$$X^* = \begin{bmatrix} x_{11} - \mu_1 & x_{12} - \mu_1 & \cdots & x_{1m} - \mu_1 \\ x_{21} - \mu_2 & x_{22} - \mu_2 & \cdots & x_{2m} - \mu_2 \\ \vdots & & & \\ x_{n1} - \mu_n & x_{n2} - \mu_n & \cdots & x_{nm} - \mu_n \end{bmatrix} \tag{3.3}$$

The sample covariance matrix

$$C = \frac{1}{n} [X^*]^T X^* \tag{3.4}$$

is then computed. Singular value decomposition is applied to the covariance matrix. C can be decomposed as follows:

$$C = UWV^T \tag{3.5}$$

where U and V are orthonormal matrices, such that $UU^T = I$, and W is a diagonal matrix. Since the covariance matrix is square, we can compute the matrix V of eigenvectors which diagonalizes the covariance matrix. In fact, for a square $(n \times n)$ covariance matrix C, there exists a rotation matrix E and a diagonal matrix Λ such that:

$$ECE^T = \Lambda \tag{3.6}$$

The principal form of C is given as:

$$\overset{n\times n}{C} = \overset{n\times n}{E\Lambda E^T} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{mn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} e_{11} & e_{21} & \cdots & e_{n1} \\ e_{12} & e_{22} & \cdots & e_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1n} & e_{2n} & \cdots & e_{nm} \end{bmatrix} \tag{3.7}$$

where columns of E and E^T are the eigenvectors and diagonal entries of Λ are the eigenvalues λ_j, *j=1,2,...n*. The sum of the eigenvalues equals the trace of the square matrix C. If the eigenvalues are sorted in descending order $(\lambda_1 \geq \lambda_2 \geq, ..., \geq \lambda_n)$, their corresponding eigenvectors, $\{e_1, e_2, ..., e_n\}$, are the principal components. Therefore matrix E contains n column vectors, representing the eigenvectors of the covariance matrix C. In general, $\lambda_i/\Sigma_i\lambda_i$ gives the relative amount of variance that the *i-th* principal component captures. The first principal component retains most of the variance, i.e. if the data are projected onto this component, more variance will be preserved than if they are projected onto any other principal component. The second component preserves the next highest residual variance, and so on. A smaller eigenvalue contributes much less weight to the total variance. In many cases, the first few components can account nearly for all the variance.

Initial Extraction of the Components

The number of components (PCs) that may be extracted from C is equal to the number of variables being analyzed. The new attributes, principal components, can be regarded as low-dimensional, more efficient representations of the same data: only the first few components will account for meaningful amounts of variance, and the later components will tend to account for only trivial variance. Therefore only the first few components will be important enough to be retained for data reduction and interpretation. An eigenvalue

(λ_j) represents the amount of variance accounted for by the corresponding component.

Determining the Number of "Meaningful" Components to Retain

It is important to determine how many components are truly meaningful and worthy of being retained for data interpretation. Three criteria may be used in making this decision: the eigenvalue-one criterion, the scree test, the proportion of variance accounted for.

- *The eigenvalue-one criterion*
  One of the most commonly used criteria for solving the "number of components" problem is the eigenvalue-one criterion, also known as the Kaiser criterion [123]. With this approach, only components with an eigenvalue greater than 1 are retained and interpreted (Fig.3.2). The rationale for this criterion is straightforward: each observed variable contributes one unit of variance to the total variance in the data set. Any component that displays an eigenvalue greater than 1 is accounting for a greater amount of variance than had been contributed by one variable. Such a component is therefore worthy of being preserved.

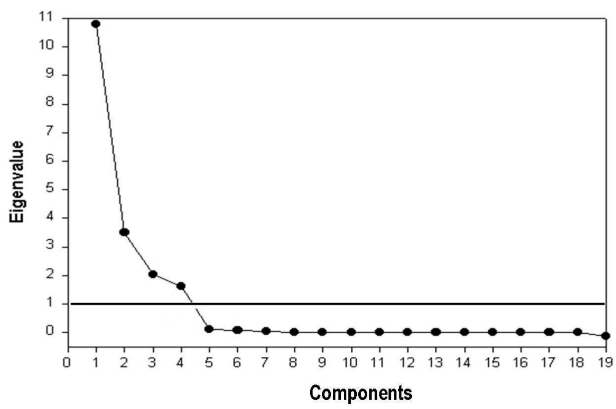


Figure 3.2: Example of the Kaiser rule for determining the number of meaningful components.

On the other hand, a component with an eigenvalue less than 1 is accounting for less variance than had been contributed by one variable. Since the purpose of principal component analysis is to reduce the number of observed variables into a relatively smaller number of components, components that account for less variance than had been contributed by individual variables are viewed as trivial, and are not retained. The eigenvalue-one criterion has a number of positive features, among which simplicity is probably the most important: no subjective decision has to be made. It has been shown that this criterion very often results in holding the meaningful number of components, particularly when a small to moderate number of variables are being analyzed and the variable commonalities are high [109]. Nevertheless there are some drawbacks associated with the eigenvalue-one criterion. Sometimes it may lead to retain the wrong number of components, especially when many variables are analyzed or when commonalities are small. In short, the eigenvalue-one criterion can be helpful when used judiciously.

- *The scree test*
  The scree test[1] was proposed by Cattell in 1966 [15]. The eigenvalues associated with every component are inserted in a bar-plot. Then a "break" is searched between the components with large eigenvalues and those with small ones (Fig.3.3).

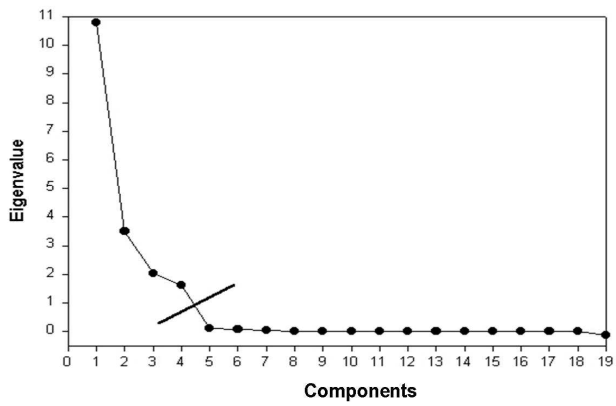


Figure 3.3: Example of the scree test for determining the number of meaningful components.

[1]The word "scree" refers to the loose rubble that lies at the base of a cliff. When performing a scree test, you normally hope that the scree plot will take the form of a cliff: at the top will be the eigenvalues for the few meaningful components, followed by a break, i.e. the edge of the cliff. At the bottom of the cliff will lie the scree: eigenvalues for the trivial components.

The break is the place where the smooth decrease of eigenvalues appears to level off to the right of the plot. The components before the break are retained; those after the break are assumed to be unimportant and are neglected. A disadvantage of this technique is that sometimes a scree plot will display several large breaks, while other times no evident break will be shown. If this happens, the choice of the meaningful components becomes subjective.

- *Proportion of variance accounted for*
  The last criterion for choosing the number of factors to be considered involves preserving a component if it accounts for a specified proportion of variance in the data set. An alternative criterion is to retain enough components so that the cumulative percent of variance accounted for is equal to a preset minimal value. In literature the threshold value for the cumulative percent of variance ranges from 70% to 90% of the overall data variability.

3.1.1 Scores and Loadings

Each principal component can be expressed as a linear combination of the original variables. The eigenvectors are weightings which, when applied to the original data, obtain principal component *scores* for the observations:

$$z = [e_{ij}]^T \cdot x_{ij}, j = 1, 2, ...h, \forall i \tag{3.8}$$

where z represents the principal component scores, e the eigenvectors and x the original data; "$\cdot$" is the dot product. Scores represent the original data mapped into the new coordinate system defined by the principal components. A large positive or negative score value indicates a variable that is correlated, either in a positive or a negative way, with the component. Referring to (3.5), the matrix L containing the data loadings is generated by:

$$L = VW \tag{3.9}$$

where V contains eigenvectors and W is a diagonal matrix. *Loadings* measure the contribution of each original variable to the principal components. Each loading is equivalent to the correlation between an original variable and a component. The strict relation between data, loadings and scores is reported in Fig.3.4.

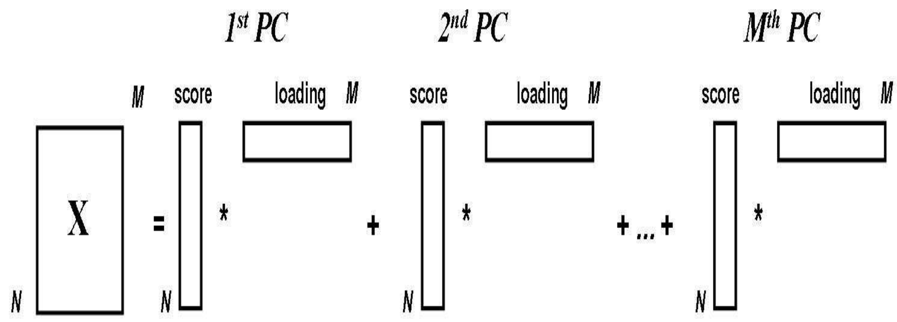


Figure 3.4: Decomposition of X matrix in its principal components basis.

3.2 Traditional PCA (*t-PCA*)

In this thesis the term "traditional principal component analysis" (*t-PCA*) is referred to the application of PCA directly to kinematic, kinetic or electromyographic data of motion analysis. Data derived from gait or other analyzed cyclic movement consist in time normalized waveforms, sampled at each 1% (1-100% of the cycle).

Each subject is described by more variables simultaneously acquired during repeated trials. Data belonging to every variable are organized in a matrix X:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,100} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,100} \\ \vdots & & & \\ x_{n,1} & x_{n,2} & \cdots & x_{n,100} \end{bmatrix} \tag{3.10}$$

where each row represents a single trial and columns contain the instantaneous values of the variable during the motion cycle.
The analytical steps described in Section 3.1 are then applied on X. First the mean is subtracted by each column, then the covariance matrix is evaluated. PCA is then applied to the covariance matrix and the principal components are obtained. Each *PC* contains 100 values, each having a factor loading. The further step is to choose the number of *PCs* to be retained. The eigenvalues or weighting factor of each *PC* indicate how many components are important in conveying most of the major information. The last step consists in giving a physical meaning to each *PC*. To facilitate interpretation of the *PCs*, it is generally suggested [93, 94, 108] to illustrate the effect of the

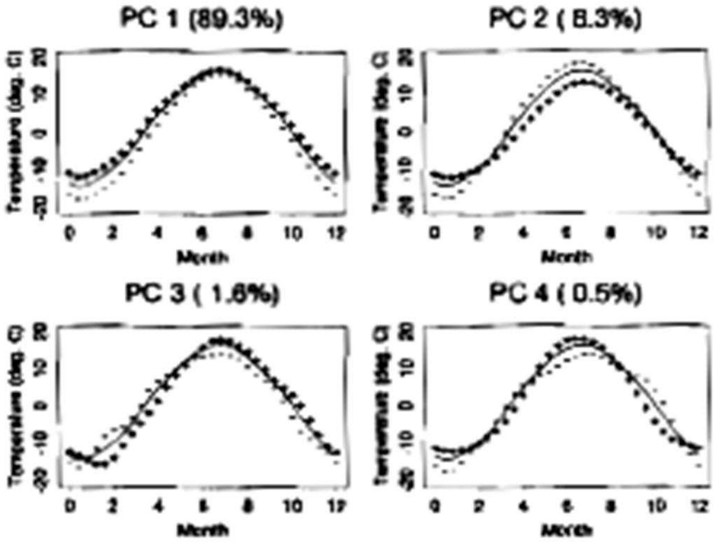


Figure 3.5: Example of the effect of principal components on the mean curve: mean curves and the effects of adding (+) and subtracting (-) a multiple of each PC curve [93].

first principal components on the mean curves. An example is reported in Fig.3.5 [93].

Hence, one method is to examine plots of the overall mean trend and the patterns obtained by adding and subtracting a suitable multiple α of the principal component in question [93, 94, 108]. The multiple α is set as:

$$\alpha = c \cdot \left[\int \{\mu(t) - \overline{\mu}\}^2 \, dt \right]^{1/2} \tag{3.11}$$

where $\mu(t)$ is the mean of the rows of *X(t)* and $\overline{\mu}$ is the average value of $\mu(t)$ over t. The factor c is a constant subjectively chosen to produce a clear visual impression of the effect of the various components.

In this way each principal component represents a specific feature of the waveform data. The principal component scores, z, measure the contribution of the principal components to each individual waveform. The percent variation explained within a feature quantifies the correlation between the variable and the feature.

In literature *t-PCA* has been applied to derive representations of gait patterns for able-body subjects. Applying PCA directly to the time series of three-dimensional positions of marker applied to the skin [26, 113] or to time series of derived joint angles [112], a low-dimensional posture model was defined and the analysis of entire gait waveforms was facilitated. Moreover *t-PCA* was used to detect the main functional structure of actions taken by specific muscles during gait [102, 103, 104]. The factor loadings within the principal components allowed to determine which gait parameters were most important in able-bodied locomotion.

Principal component modeling of gait kinematic and kinetic parameters reduces the data to measures of distance from normal. These measures are shown to be sensitive to changes in gait pattern associated with traumas or injuries. They allow the individualization and quantification of "disability signs" not perceived by visual inspection. t-PCA was adopted to provide insight into gait data obtained from persons with stroke [82] or subjects affected by osteoarthritis [31, 32, 33]. This technique permitted to quantitatively characterize the gait of pathological subjects with respect to subjects whose gait was considered to be normal. In a recent work *t-PCA* was simultaneously applied to waveform and discrete measures [2]. Waveform measures were dynamic gait measurements that vary continuously in time throughout gait, while discrete measures were single values related to stride characteristics or clinical parameters. The results of the study emphasized the ability of *t-PCA* in representing an objective method for reducing and analyzing interrelated gait measures.

Possible applications of *t-PCA* includes also the analysis of multi-joint coordination for upper [42] and lower limbs [23] during cyclical movements. The application means to account for the relations among multiple variables. It allows to discover the spatio-temporal structure of coordination among body segments, thus providing a global measure of performance.

3.3 Functional PCA (*f-PCA*)

The basic philosophy of functional data analysis is to think each function fitted to a set of data as a single observation. Actually, data are supposed to have an underlying functional relationship governing them. It has the advantage that no assumption is made about the function or the data. The term "functional" refers to the intrinsic structure of the data rather than to their explicit form [93]. Main goals of functional data analysis are:

- to represent the data in ways that aid further analysis,
- to display the data so as to highlight various characteristics,

- to study important sources of pattern and variation among data,
- to compare two or more sets of data with respect to certain types of variation, where two sets of data can contain different sets of replicates of the same function or different functions for a common set of replicates.

Functional data analysis can be divided in a sequence of simple steps. The first task is estimating smooth functions from discrete noisy data.

Smoothing

A kinematic, kinetic or electromyographic variable acquired in the *i-th* trial (y_i) consists in a set of n discrete measured values $y_{i1}, y_{i2}, ..., y_{in}$. These values are converted to a function $x_i(t)$ computable for any desired argument value t through smoothness. Smoothness is a property of x and may not be obvious in the raw data vector y because of observational error or noise that may be imposed on the underlying signal. In other terms, it can be written:

$$y_j = x(t_j) + \epsilon_j \tag{3.12}$$

where the error ϵ_j contributes a roughness to the raw data. One of the tasks in representing the raw data as functions is to filter out this noise as efficiently as possible. Moreover, the natural biovariability present in multiple repetitions of the same movement has to be taken into account. The standard statistical model for the ϵ_j is that they are independently distributed with zero mean and finite variance.

One of the best known basis expansion for smoothness is the β-splines smoothing. Differently from Fourier and polynomial bases, this method enables to fit local features in the curves. A β-spline consists of a set of fixed positions, called “knots”, and piecewise smooth curves, called “basis functions”, connecting each of the knot positions, satisfying continuity conditions between the pieces.

A simple example of β-splines [37] is shown in Fig.3.6. A β-spline of degree 1 (Fig.3.6(a)) consists of two linear pieces, one from x_1 to x_2, the other from x_2 to x_3. A β-spline of degree 2 (Fig.3.6(b)) consists of three quadratic pieces, joined at two knots. At the joining points ordinates of the polynomial pieces and their first derivatives are equal.

The β-splines overlap with each other: first degree β-splines overlap with two neighbors, second-degree β-splines with four neighbors (right side of Fig.3.6). It is possible to construct a large set of β-splines, by introducing more knots. The general properties of a β-spline of degree q are:

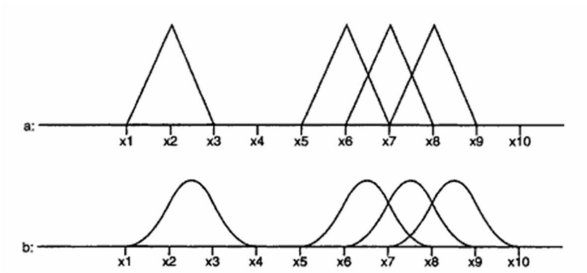


Figure 3.6: One isolated β-spline and several overlapping ones with: (a) degree 1, (b) degree 2 [37].

- it consists of *q+1* polynomial pieces, each of degree q;
- the polynomial pieces join at q inner knots;
- at the joining points, derivatives up to order *q-1* are continuous;
- the β-spline is positive on a domain spanned by *q+2* knots, everywhere else it is zero;
- except at the boundaries, it overlaps with *2q* polynomial pieces of its neighbors.

β-splines can track local shape deformations using a small number of parameters, unlike Fourier descriptors which require many parameters and can have spurious oscillations. β-splines give complete control over flexibility, allowing more flexibility where needed and less where not needed. A demonstration of β-spline flexibility can be inferred by Fig.3.7.

Let equidistant knots be considered and let $B_j(x,q)$ denote the value at point x of the *j-th* β-spline of degree q for a given equidistant grid of knots. A fitted curve y to data *(x,y)* is the linear combination:

$$y(x) = \sum_{j=1}^{n} a_j B_j(x,q) \tag{3.13}$$

where n is the number of used β-splines.

Penalty

The smooth functions related to data are obtained using a least square approach. Considering the regression of m data points (x_i, y_i), *i=1, 2,..., m* on

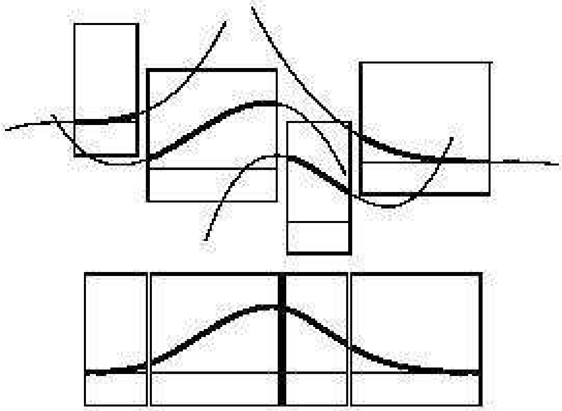


Figure 3.7: A simple example for a β-spline of order 3. The four cubic polynomials, included in rectangles in the upper part of the figure, are joined at the three knots, to form the curve represented in the lower part.

a set of n β-splines B_j, *j=1, 2,..., n*, the least square function to minimize is:

$$S = \sum_{i=1}^{m} \left\{ y_i - \sum_{j=1}^{n} a_j B_j(x_i) \right\}^2 \tag{3.14}$$

where a_j are suitable coefficients.

If the number of knots is large, the fitted curve will show more variation than that justified by the data. A penalty can be introduced to make the result less flexible [85]. The extra roughness penalty term, controlled by a smoothing parameter λ, ensures that each fitted curve is determined not only by its goodness of fit but also by its roughness. Roughness is measured by integrating the second derivative of the fitted curve:

$$R(x) = \int_{x_{min}}^{x_{max}} \left\{ \sum_{j=1}^{n} a_j B_j''(x, q) \right\}^2 dx \tag{3.15}$$

where x_{min} and x_{max} are the limits of the curve. A compromise is then struck between fitting the data and keeping the fit smooth, forming the objective function:

$$S = \sum_{i=1}^{m} \left\{ y_i - \sum_{j=1}^{n} a_j B_j(x_i) \right\}^2 + \lambda \int_{x_{min}}^{x_{max}} \left\{ \sum_{j=1}^{n} a_j B_j''(x, q) \right\}^2 dx \qquad (3.16)$$

When λ=0, only fitting the data matters; as λ increases, more emphasis is placed on penalizing roughness; when λ is very high, only roughness matters and functions having zero roughness are used. There are data-driven methods for choosing λ, but smoothing inevitably involves judgment. Usually [93, 94, 108] cross-validation is adopted to determine a starting point for possible values of λ before a subjective choice is made.

PCA for functional data

The definition of principal component analysis on functional data is analogue to the traditional PCA's one. Assuming that observed data are $x_i(t)$ in the functional situation, the covariance function C is given by:

$$C(s, t) = \frac{1}{N} \sum_{i=1}^{N} x_i(s) x_i(t) \qquad (3.17)$$

where s and t are instants of time, N is the number of functions. Thus the functional PCA problem leads to the eigen-equation:

$$\int C(s, t)\ \xi(t) dt = \rho\ \xi(s) \qquad (3.18)$$

where ξ are the eigenvectors and ρ the eigenvalues. Thus the first principal component is the weight function $\xi_1(s)$ for which the principal component scores

$$f_{i1} = \int \xi_1(s)\ x_i(s) ds \qquad (3.19)$$

maximize $\sum_i f_{i1}^2$ subject to

$$\int \xi_1^2(s) ds = \|\xi_1\|^2 = 1. \qquad (3.20)$$

The second principal component $\xi_2(s)$ is the weight function for which scores maximize $\sum_i f_{i2}^2$ subject to $\|\xi_2\|^2 = 1$ and the additional constraint

$$\int \xi_2(s)\ \xi_1(s) ds = 0 \qquad (3.21)$$

and so on.

PCA can be used to study the simultaneous variation of more than one function. Bivariate functional PCA (*bf-PCA*) is simply the extension of *f-PCA* to deal with bivariate functional data. The analysis is carried out by replacing two variables with a suitable β-splines expansion. The two vectors containing the coefficients are then concatenated in a single long vector. The covariance matrix is then evaluated. Further analytical steps are the same as previously described for *f-PCA*.

Ramsay and Silverman [93, 94, 95] were important supporters of the application of functional data analysis to gait data. They offered a complete and useful reference to the main concerns of functional principal components analysis (*f-PCA*) and showed many examples of how this technique works in different disciplines including biomechanics.

This technique has been often adopted for the tracking of human motion in video sequences [83, 84]. It is a fully automated method for learning periodic human motions from training data. Concerning sports biomechanics, Ryan [100] and Harrison [51] applied *f-PCA* to knee sagittal angle kinematics during the vertical jump for a group of young boys. These works demonstrated the potential benefits of functional data analysis in providing greater insight into the changes of kinematic patterns.

3.4 Two-stage PCA (*2-PCA*)

A useful technique for gait recognition from motion capture has been proposed by Das [29, 30]. It consists in a two-stage principal component analysis. The first step provides a low dimensional representation of gait or cyclic movement, the second one enables to distinguish clusters corresponding to the individual identity or to the type of motion.

First stage

Two-stage principal component analysis is used to analyze kinematic data of more trials of the same subject or of different subjects. It is simultaneously applied on more variables. An n-dimensional data-set is focused on n variables, usually joint angles and angular velocities. Thus each data point is a n-dimensional vector consisting of values x_{ij} of variable *i=1, 2,..., n* at time *j=1, 2,..., 100* (percentage of the movement cycle). If each such data point is denoted by $X_j = [x_1, x_2, ..., x_n]$ and the principal components are denoted by P_k, (*k =1, 2,...,n*), then the projections Y_{kj} of X_j on P_k are given by:

$$Y_{kj} = \langle X_j, P_k \rangle \tag{3.22}$$

where $\langle\ \rangle$ represents the dot product. For every movement cycle (100 percentage segment) P_k and Y_{kj} are calculated separately. Then the projection of the original data onto the first two principal components (Y_{1j} and Y_{2j}) are considered. The plot of the second principal component versus the first one consists in a manifold similar to the one represented in Fig.3.8.

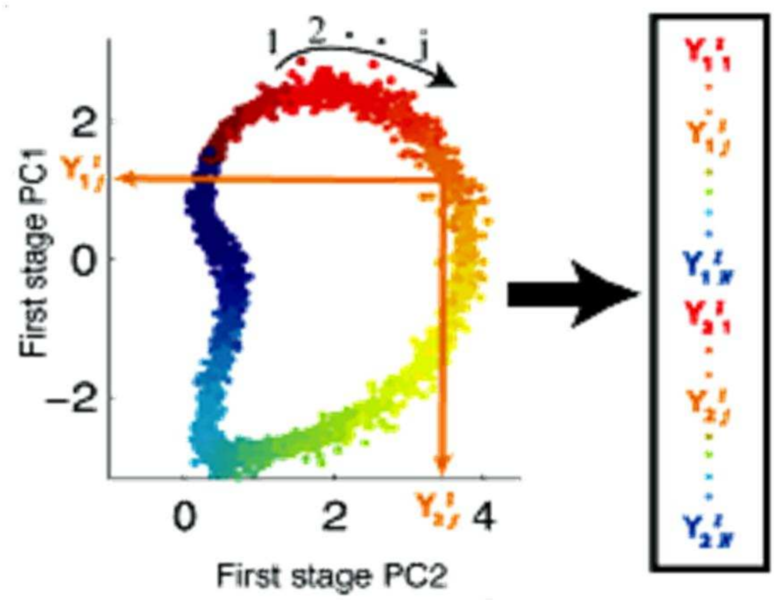


Figure 3.8: Construction of each data point for the second PCA decomposition. The manifold plotted here represents the projection of data points X_j on the first two components. The colors represent the phase of gait. For each gait cycle divided into N normalized time points, the projection values $Y2_j$ are plotted versus $Y1_j$ [30].

Therefore, the variation in the manifold and the temporal dynamics of how the manifold is traversed during a movement cycle, contain information about differences in trials of the same subject or in trials of different subjects.

Second stage

In the second stage of *2-PCA*, the information about the temporal variability of the data throughout the movement is captured. The projection values Y_1 and Y_2 for each movement cycle are concatenated in a single time series. If the *l-th* movement cycle is represented by N uniformly spaced time points, each data point Y^l, derived from the first step of *2-PCA*, is given by:

$$Y^l = \begin{bmatrix} Y_1^l & Y_2^l \end{bmatrix} \tag{3.23}$$

The second principal component decomposition is then performed on each cycle. The projection Z^{lm} of Y^l on the principal component Q^m of the second stage is given by:

$$Z^{lm} = \left\langle Y^l, Q^m \right\rangle = \left\langle Y_1^l, Q_1^m \right\rangle + \left\langle Y_2^l, Q_2^m \right\rangle \tag{3.24}$$

where $\langle \; , \; \rangle$ represents the dot product, m= *1, 2,..., 2N* and Q_1^m and Q_2^m are the part of component Q^m that act respectively on the first and second dimensions.

The second stage of *2-PCA* extracts the information about the type of motion, the identity of the subject, etc. contained in the manifold. Projections in the eigen-space after two-stage PCA are used to perform various classification tasks. An example is reported in Fig.3.9, which shows results of identity classification on six running subjects [30].

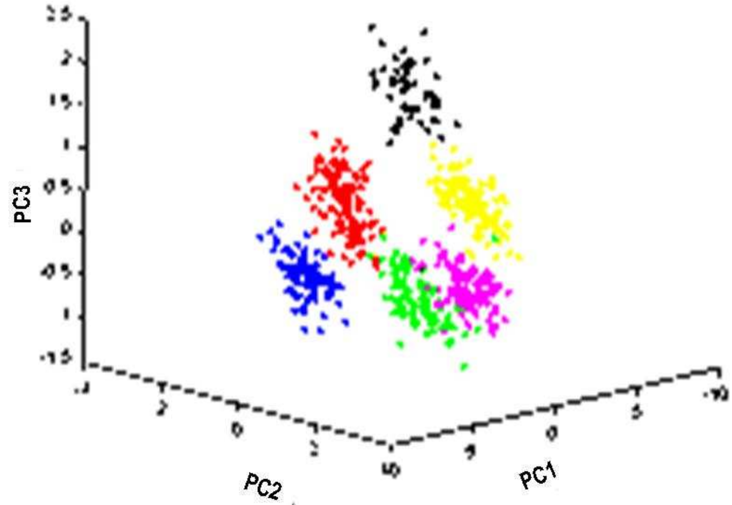


Figure 3.9: Recognition of different people from their running gaits. Points corresponding to gait cycles of each individual are color coded in the second-stage PC space. Points for each individual are distinguishably clustered [30].

Das introduced the two-stage PCA as a tool to classify gait (run, jog, etc.) and to recognize individuals [29, 30]. In psychophysical studies it has been found that observers are sensitive to specific "motion features" that characterize the relative motion of body parts. Thus the main idea is to identify a low-dimensional manifold of motion parameters described by orientation angles of limb segments. These manifolds could be sensitive descriptors of gait in biomechanics.

Das found that the first two principal components of gait data had close correspondence with the identified motion features. A description of deviations in the manifold across individuals or types of gaits was also performed, taking into account the temporal characteristics of the motion. The *2-PCA* technique demonstrated high accuracy for gait data and identity discrimination tasks, and revealed to be a good tool for getting a biomechanical interpretation of individual movements.

3.5 Concluding remarks

PCA is a key technique to consider in motion analysis. First, it allows to reveal new and interesting aspects of data, as well as to highlight already known features within them. An indication of the complexity of the data is also provided, in the sense of how many types of curves and characteristics are to be found.

A close relation exists between *f-PCA* and *t-PCA*. A significant, intrinsic difference between the two methods lies in the perception that functional data are observed in the continuum, without noise, whereas traditional data are observed at distributed time points and are often subject to experimental error [46]. Differences between the two techniques relate to the way in which a problem is perceived and are more conceptual than actual.

t-PCA consists in the application of PCA to data described on a dense temporal grid. The implementation is very simple, since only the routines of the multivariate PCA are needed. However, the eigenfunctions are rough (non-smooth). When data consist in curves represented only on a small number of distributed points, non-negligible measurement error is present. In this case, *t-PCA* presents bad results, especially when data curves do not have similar shapes, or are so sparse that the individual data profiles cannot be discerned. Measurement errors can also mask the shapes of the underlying profiles. This motivates the application of functional data approaches, and in particular, functional principal component analysis, to this kind of data.

f-PCA is a statistical method that corresponds more to the functional nature of data. When the data are observed with experimental error, the operation of smoothing data can greatly reduce the effects of noise. It has been shown [46] that estimation of eigenvalues can be consistent, even if only a few observations are made of each function, and each observation is encumbered by noise. Nevertheless, application of *f-PCA* implies a further effort in searching an optimal method to smooth the eigenfunctions. In fact, smoothing requires some care in respecting the periodicity and the nature of data. *f-PCA* has also the advantage to get more stable and better interpretable results.

The *2-PCA* approach is a powerful tool in identifying a set of physical features for the recognition of the motion of different subjects. The manifold representation normalizes the range of original angle variables as well as temporal evolution of components in a motion cycle [30]. This makes the representation time invariant but discards potentially discriminating information.

Chapter 4

Materials and Methods

This chapter contains the information about the experimental protocol and data processing. A dedicated is reserved to the preprocessing of data, in order to facilitate the interpretation of results.

4.1 Subjects

Four male (weight 64 $\pm$ 2.4 kg, height 180.8 $\pm$ 9.1 cm) and three female (weight 50.7 $\pm$ 6.8 kg, height 167.3 $\pm$ 5 cm) race walkers of national and international class were recruited for this study. All the athletes were training at least 6 training sessions a week. Written informed consent was obtained from all the subjects.

Race walkers were numbered in decreasing order according to their skill levels and named "$s1$", "$s2$", etc. The ordering was carried out by considering their competition results and trainers information. Race walkers were divided in three groups: upper (group1), intermediate (group2), lower (group3) levels. Group1 contained the first three subjects: $\{s1, s2, s3\}$; $\{s4, s5\}$ were assigned to group2; $\{s6, s7\}$ to group3. The number of race walking repetitions for each class varied, with 58 trials in group1, 40 in group2 and 58 in group3. Detailed information about the anthropometric characteristics and the agonistic results are reported in Tables 4.1 and 4.2 respectively.

All the athletes were selected taking care they did not show any remarkable lower limb injury or dysfunction at the time of the experiments.

		s1	s2	s3	s4	s5	s6	s7	μ	σ
sex	[m,f]	m	f	m	m	f	m	f		
age	[years]	22	20	18	23	18	19	18	19.7	2.1
height	[m]	1.74	1.62	1.90	1.87	1.68	1.72	1.72	1.75	0.1
weight	[kg]	64.0	43.0	61.0	67.0	53.0	64.0	56.0	58.3	8.3

Table 4.1: Anthropometric data for subjects under analysis

	5 km		10 km		20 km	
	t	$\overline{v}$	t	$\overline{v}$	t	$\overline{v}$
s1	19:58.00	4.17	40:56.74	4.07	1:25:39.0	3.89
s2	22:55.20	3.64	46:38.53	3.57	-	-
s3	21:03.66	3.96	42:22.59	3.93	-	-
s4	20:06.61	4.14	42:59.95	3.88	-	-
s5	23:25.60	3.56	48:34.43	3.43	1:39:47.0	3.34
s6	21:56.33	3.80	44:24.97	3.75	1:33:06.0	3.58
s7	24:04.61	3.46	-	-	-	-
μ		3.82		3.77		3.60
σ		0.28		0.24		0.28

Table 4.2: Athletes' personal best over the most common distances of race walking competitions. The performances achieved over the 5, 10 and 20 km events are reported in the following format: h:mm:ss, where h stands for hours, m for minutes and s for seconds. Dashes mean that the athlete did not compete over that distance. $\overline{v}$ represents the average progression speed, in m/s. μ is the mean value and σ is the standard deviation.

4.2 Instrumentation

Kinetics and kinematics data were acquired at the Laboratory of the Department of Bioengineering of the Politecnico di Milano[1].
A stereophotogrammetric system (ELITE 2002, BTS, Milan, Italy) composed by 8 cameras working at 100Hz was used to acquire the race walking movement. Cameras were set in order to enable athletes to perform their movement in a calibrated volume of about 8×2×4 meters.

The disposition of cameras, shown in Figure 4.1, allowed to detect the position of markers placed on both sides of subjects throughout the race walking movement.

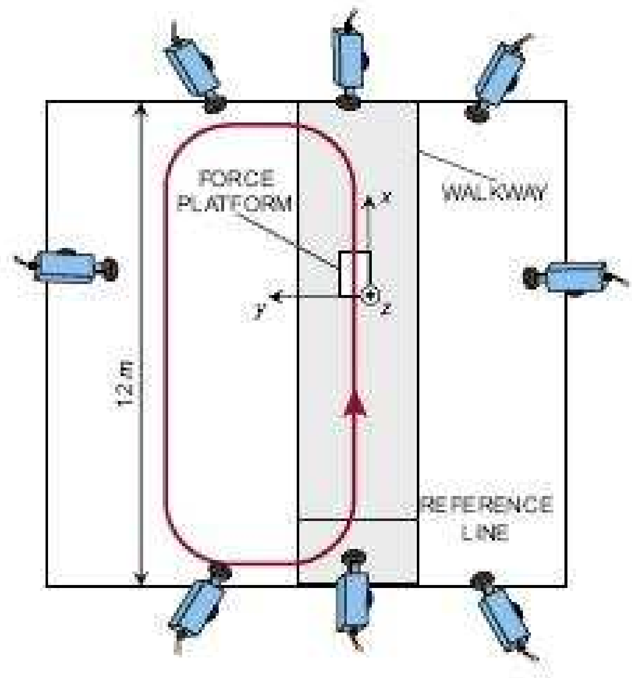


Figure 4.1: Experimental Setup [90]

Before each experimental session the system was calibrated and a maximum mean error of 1.5mm (concerning the length of a 600mm rigid bar) was tolerated. A force plate (AMTI OR6-7-1000, Watertown, USA) with a sampling rate of 500Hz was used to capture the ground reaction force (GRF). The reference frames relative to kinematics and ground reaction force were

[1]All experimental session were carried out at Laboratorio di Analisi della Postura e del Movimento "Luigi Divieti", Department of Bioengineering, Politecnico di Milano, Italy.

different. Therefore, they were adapted to a common convention: the x axis corresponded to the antero-posterior axis, oriented to the direction of progression; the z axis corresponded to the vertical direction; the y axis corresponded to the medio-lateral direction (cross-product of x and z).

4.3 Marker-set

Some preliminary tests were performed at the Laboratory. Their purpose was the comparison of two protocols (Davis and SAFLo[2]) [90], to define the most suitable one to reliably detect lower limb joint angles on the sagittal plane, without altering natural movement. Both marker-sets are reported in Figures 4.2 and 4.3.

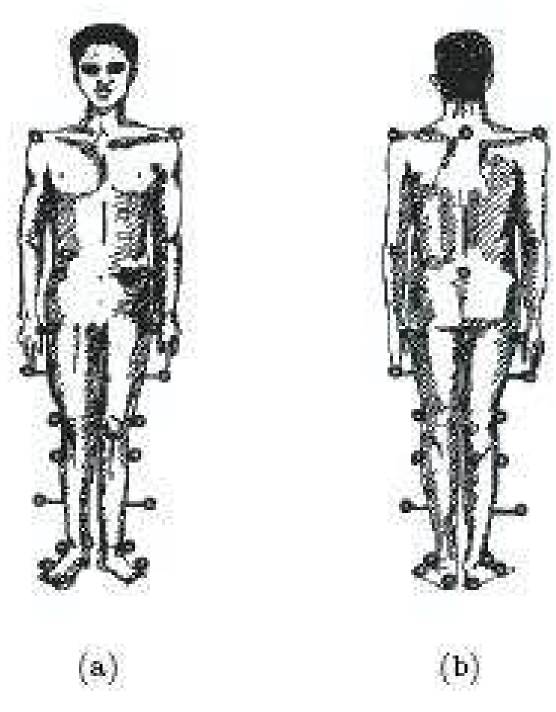


Figure 4.2: The Davis marker set [90]. Front view (a) and back view (b).

The ground reaction force during heel strike is significantly greater during race walking than in normal walking [13]. This is related both to the increase

[2]Servizio di Analisi della Funzionalità Locomotoria.

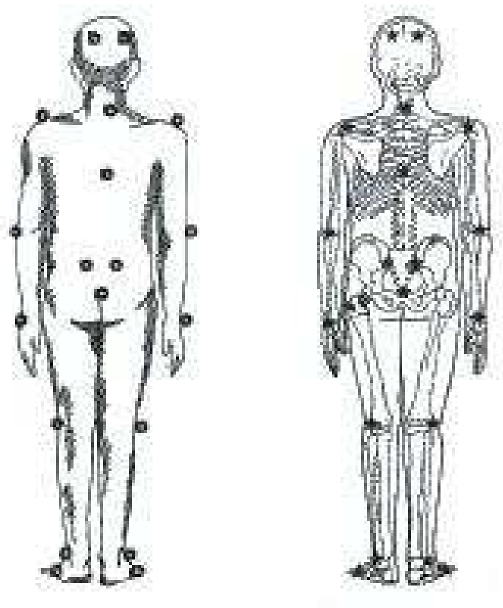


Figure 4.3: The SAFLo marker set [90]. All markers are visible from the back view.

in walking velocity and to the full extension of the knee required by the rules. Davis configuration involves the application of markers placed on bars, which feel more the effect of skin artifact during such a fast movement. The oscillation of markers has repercussions on joint angle reconstruction, considerably increasing data noise. Moreover the markers placed on the greater trochanter did not allow athletes to wave the arms close to the trunk, thus challenging the natural movement. SAFLo configuration has the disadvantage to have only two markers placed on each segment of the lower limbs, thus not providing any information about thigh, leg and foot rotations. Nevertheless, this marker-set has the advantage not to have wands encumbering the subject, vibrating and moving relative to the underlying skeleton. Therefore the SAFLo protocol was adopted. It consists of a total-body marker-set, with 19 retroreflective hemisferical markers (15mm diameter) fixed on specific anatomical landmarks: lower prominence of the sacrum, posterior superior iliac spines, lateral femoral condyles, lateral malleoli, fifth metatarsal heads (for the pelvis and lower limbs section); seventh cervical vertebra and point of maximum kyphosis (for the column); acromion bones, lateral homerus epicondyles and stiloideus processes (for the upper limbs section); parieto-occipital areas of the head. Markers were glued to the skin with care, not to be detached or moved because of rapid movements or sweat.

4.4 Data collection

Race walkers performed a standard 20 minutes warm up routine and some trials to familiarize with the experimental settings. The 15m long walkway allowed them to perform a technically correct race walking, with an approximately constant speed. The position of the force platform at two-thirds of the path, let the athletes accelerate and reach a stable progression before getting to it. Only trials in which they randomly put their left or right foot on it were recorded. As many as 20 suitable race walking trials (for left and right side) were acquired for each athlete. They were performed at a self-selected pace and under the trainer supervision. The trainer checked the goodness and correctness of race walkers performance. Four subjects accepted to participate to two testing sessions (called session I and II) in different periods of the agonistic season, according to their athletic engagements.

4.5 Data processing

Despite of acquiring only technically correct race walking performances, with the foot completely inside the surface of the force plate, a further control was applied on the performed trials. The complete description of data preprocessing is reported in [91]. A first monitoring of movement correctness derived from the evaluation of the ground reaction force pattern. Since maintaining an almost constant speed implies similarity between breaking and propulsive impulse of GRF, the anterior and posterior force areas were compared. Moreover a post-processing control directly was applied on the progression speed of the center of mass, comparing its values at heel strike and toe-off with the usual training speed of the athlete. From total body kinematics, the position, velocity and acceleration of the center of mass (COM) were carried out. Three dimensional coordinates of internal joint centers, joint angles, velocities and accelerations were estimated from anthropometric measures through specially designed algorithms [27, 88]. Hip, knee and ankle joint moments were measured by using the Newton-Euler free body dynamic equilibrium equations. Each body segment mass, inertial moments, and gravity center positions were estimated through Zatsiorsky and Seluyanov regression equations [125]. Concerning sign conventions, lower limb joint extension moments were defined as positive. Ground reaction force was normalized by body weight; moments and powers were normalized by body weight and height [55, 76].

The analysis was focused mainly on 15 variables, supposed to be the most consistent measures of lower limb motion analysis [24, 25, 41, 45, 61, 62, 92, 116]:

- antero-posterior, mediolateral and vertical ground reaction forces ($R_{ap}(t)$, $R_{ml}(t)$, $R_v(t)$);
- hip, knee and ankle joint angles (Fig.4.4) in the sagittal plane ($A_{hs}(t)$, $A_{ks}(t)$, $A_{as}(t)$);
- hip, knee and ankle joint angular velocities in the sagittal plane ($V_{hs}(t)$, $V_{ks}(t)$, $V_{as}(t)$);
- hip, knee and ankle joint moments in the sagittal plane ($M_{hs}(t)$, $M_{ks}(t)$, $M_{as}(t)$);
- pelvic tilt, pelvic obliquity and pelvic rotation angles ($A_{pt}(t)$, $A_{po}(t)$, $A_{pr}(t)$).

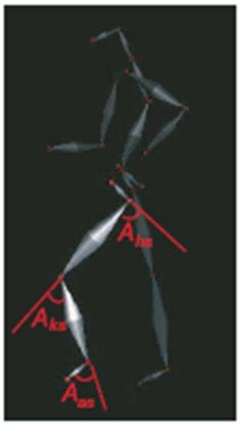


Figure 4.4: Conventions for angles definition in lower limb joints. Average values from the whole population, during standing, were: $A_{hs} \cong 25deg$, $A_{ks} \cong 5deg$, $A_{as} \cong 70deg$ [90].

The race walking cycle of every acquisition was analyzed. In classical clinical gait analysis the "gait cycle" is the period of time from the heel strike of one foot to the following heel strike of the same foot. For this study the "race walking" cycle was considered as the interval of time from the toe-off to the following toe-off of the same foot. This choice derived from the need to optimize the quality of data; in fact, when the first foot centered the force plate, the second heel strike happened out of the acquisition volume. Therefore in the identification of the race walking cycle, the first toe-off was before arriving at the force platform. It was estimated by using specially designed algorithms that look at the kinematic variables, while the second one came from ground reaction force measure. The first toe-off corresponded to the peak of vertical acceleration of the marker applied to the fifth metatarsal of the supporting leg. The second one corresponded to the instant in which $R_v(t)$ reached the base line.

After extracting the race walking cycle from the selected variables of each repetition, unrepresentative trials were removed thanks to a double control. First macroscopically anomalous curves were eliminated; then the cycle duration was checked, in order to remove temporal outliers. A threshold of 1.5 interquartile from the sample median time value was chosen in agreement with Chau [19].

Curve registration

Data were normalized, resampling the race walking cycle time to 100 percentage frames. This procedure, commonly adopted in literature, is useful in the study of inter-subject evaluations. However, since this approach is a form of artificial alignment of data, it may mathematically separate data from their functional principal components. This was one of the problems encountered by some researchers in their studies [21, 119, 120].

Curve registration [93, 96] was applied to data to evaluate the effect of time domain variability. Registration [96] is the process of temporally aligning a set of curves, by minimizing discrepancies from an iteratively estimated sample mean or by aligning specific curve landmarks. Landmark registration has been adopted in literature to reduce inter-subject variability in angular displacement, moment and power curves [19, 101, 102, 104] or to analyze phase variability in gait curves [96, 108]. In this study a minimum variation was measured among registered and not registered data. Curve registration presented no major effect, probably because anomalous and temporally odd curves have already been removed. This proof of data consistency induced the choice of adopting normalization on unregistered data.

4.5.1 Data arrangement for *t-PCA* and *f-PCA*

For the univariate application of *t-PCA*, each subject was represented by a set of data curves arranged in a three-dimensional matrix. Rows corresponded to repeated trials, columns to percentage frames of the race walking cycle (1-100%), while the third dimension stored the 15 analyzed variables. The matrices of all the race walkers were vertically concatenated (Fig.4.5).

Therefore all the trials of the seven athletes were grouped in a $n \times 100 \times 15$ matrix. The number of rows corresponded to the sum of trials of all the athletes, columns to percentage frames of race walking cycle and layers to variables.

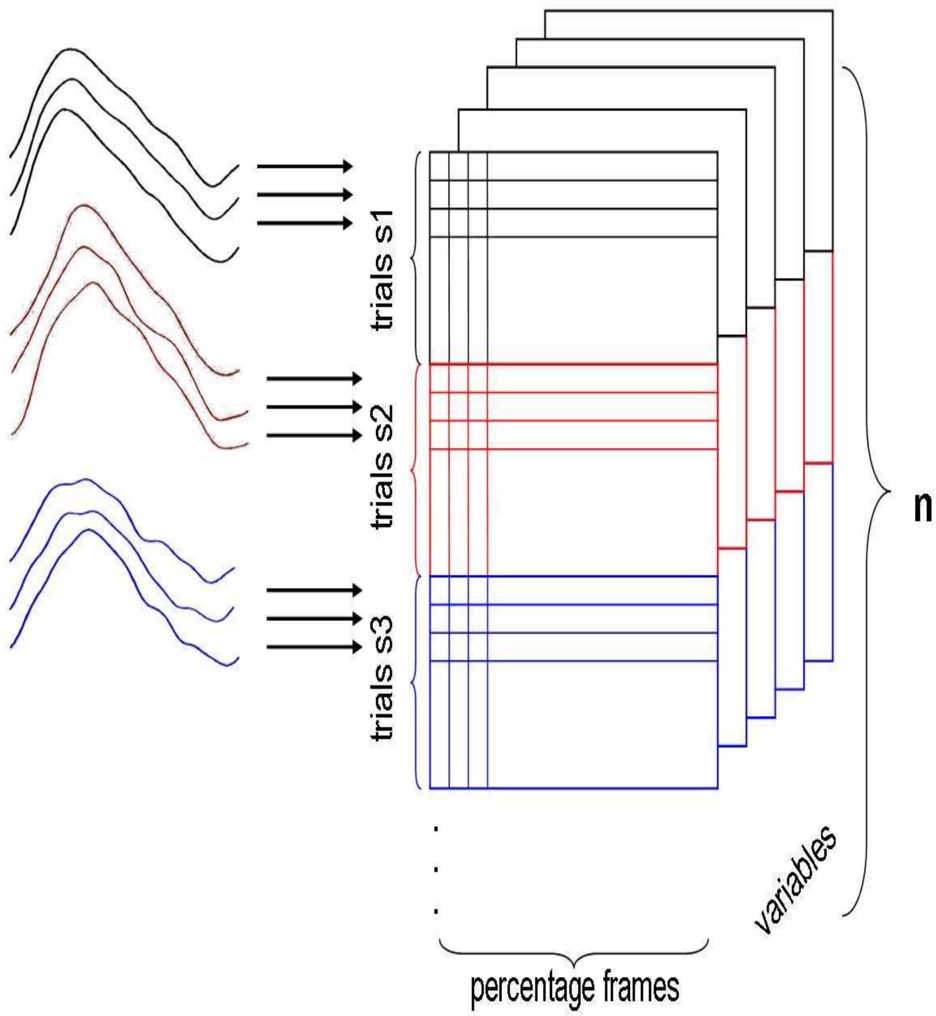


Figure 4.5: Example of the data arrangement for athletes' analysis through univariate and bivariate *t-PCA* and *f-PCA*.

An analogous arrangement was adopted for *f-PCA*, but time series were first smoothed through β-splines and they were then inserted in the rows of the matrix shown in Fig.4.5. PCA was applied to each variable, coinciding with a layer. Principal components explaining a total amount of 95% of data variance were inserted in a bar plot. The scree test was then applied to choose how many components to retain. They were named "*t-PCs*" and "*f-PCs*" respectively for the two techniques.

A scatterplot analysis was carried out, to better inspect the ability of traditional and functional PC scores to differentiate athletes according to their skill level. It consisted in a graphical view of the scores on the first and second principal components. Points corresponding to trials of the same subject were shown with the same color. In the meantime a biomechanical interpretation was derived by the evaluation of the deviation from the mean caused by these *PCs* (Paragraph3.1). It consisted in the plot of the mean curve of the variable, with curves created by adding and subtracting a multiple of the principal component.

For coordinative inspection bivariate traditional (*bt-PCA*) and funtional (*bf-PCA*) PCA were used. Data were arranged as previously described in an $n \times 100 \times 15$ matrix. Two variables (coinciding with two layers) were taken into account simultaneously. This technique allowed to inspect the coordination between angles of consecutive joints: hip and knee or knee and ankle.

Finally, boxplots were used to compare functional principal component scores of session *I* respect to session *II*. This statistic tool produces a box with lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers. Points beyond the whiskers are displayed using "+".

Preprocessing for *2-PCA*

For two stage PCA (*2-PCA*) data were arranged differently. In the first stage, PCA was simultaneously applied to more variables. Thus each subject was described by a matrix which rows corresponded to percentage frame (100), columns to variables (15) and layers to the n repeated trials (Fig.4.6).

Principal component analysis was then applied to this matrix and the first two extracted principal components (*2-PC1*, *2-PC2*) were vertically concatenated to form a 200-frame time series for every trial. In the second stage, principal component analysis was applied to these curves.

Loadings of the first stage *2-PCA* were inserted in a bar plot. They represented the relative contribution of each variable to the first and second principal components. The scores for *2-PC1* and *2-PC2* were inserted in a scatterplot, to form the "O"-shaped manifold described in Paragraph3.4.

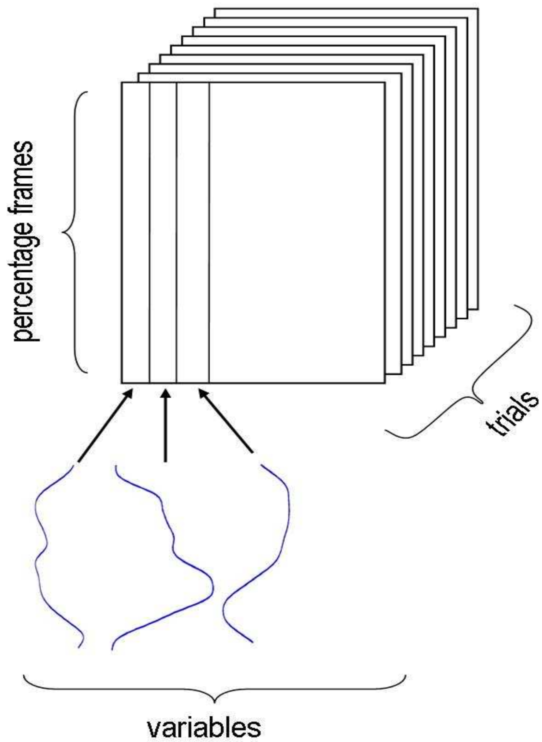


Figure 4.6: Example of the data matrix for every subject in the first stage of *2-PCA*.

Frames associated to different phases of the race walking cycle had different colors: red for the swing phase (SP), green for the frontal leg support phase (FSP) and cyan for the rear foot support phase (RSP). Also the scores of the second stage *2-PCA* were inserted in a scatterplot. In this case different colors were associated to different athletes.

Chapter 5

Results and Discussion

Since all the athletes were race walkers of high level, differences in movement performances were not easy to be recognized by simple inspection of kinematic or kinetic variables. An example concerning the knee joint angle and the ankle moment is reported in figure 5.

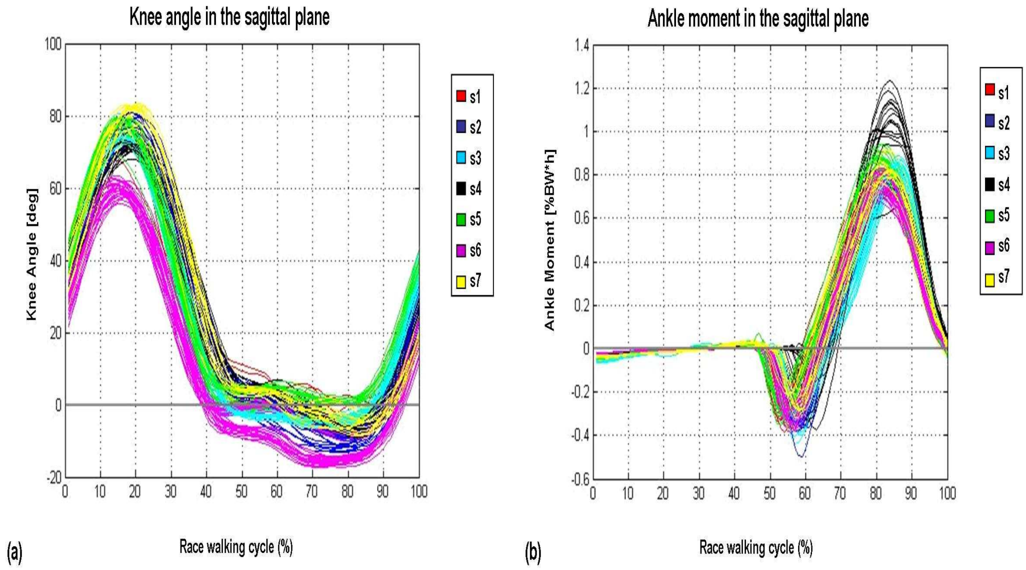


Figure 5.1: Knee angle (a) and ankle moment(b) patterns for the seven race walkers.

Each subject was represented with a different color. For each athlete all the repeated trials were plotted; time was reported as a percentage of the race walking cycle. Considering the knee joint angle (Fig.5(a)), except for *s6*, who showed a very low knee flexion throughout the movement, the other athletes displayed similar kinematic trends at the knee. On the contrary, only *s4* presented an immediately recognizable pattern (Fig.5(b)) for the ankle moment. Thus, if inspecting distinct variables, athletes might be grouped in different ways, according to their similarities and differences in variables' patterns. PCA allowed to go deeper in the analysis of individual skill characteristics.

5.1 *t-PCA*, *f-PCA* comparison

First, each variable was analyzed separately to describe the main biomechanical characteristics of race walkers. Both *t-PCA* and *f-PCA* techniques were applied to kinematic and kinetic data.

A really small number of extracted features explained greater than 95% of the variation in the data for most of the variables (Table5.1).

	t-PCA		*f-PCA*	
Variable	Var expl (%)	n. of *PCs*	Var expl (%)	n. of *PCs*
$R_{ap}(t)$	95.00	5	95.55	4
$R_{ml}(t)$	96.15	5	96.65	4
$R_v(t)$	97.22	5	96.58	4
$A_{hs}(t)$	95.49	6	95.97	6
$A_{ks}(t)$	96.13	4	96.31	4
$A_{as}(t)$	96.20	6	96.36	6
$M_{hs}(t)$	95.68	8	95.96	6
$M_{ks}(t)$	96.20	7	97.38	6
$M_{as}(t)$	96.32	5	95.40	4
$P_{hs}(t)$	95.94	12	95.32	9
$P_{ks}(t)$	95.19	5	96.11	5
$P_{as}(t)$	95.24	5	95.19	4
$A_{pt}(t)$	95.60	7	95.11	6
$A_{po}(t)$	95.17	7	95.71	7
$A_{pr}(t)$	95.85	6	96.07	6

Table 5.1: Principal component models: percentage of total explained variance and number of components

In the comparison between *t-PCA* and *f-PCA*, the first technique needed a larger or equal number of components to explain the same amount of variability. Moreover, when the scree test was applied, *f-PCA* presented a more clear-cut division between the variables to be retained and the negligible ones.

This could be due to the more sensitivity of *t-PCA* to noise. In fact, data are acquired with experimental error. Roughness of data might induce to retain also some noise, considering it information. Instead, for *f-PCA* the operation of smoothing can reduce the effects of noise. These results are in agreement with [46], who found that even in the presence of noise, statistical smoothing successfully exploits the high-dimensional character of the data.

Scores obtained from the two statistical techniques were similar for the first principal components. Moreover, the representation of the influence of the *PCs* on the mean trend was quite the same for *t-PCA* and *f-PCA*.

These findings led to the hypothesis that kinematic and kinetic data contained an inborn factor which could be revealed by both the analyses. This was probably due to the intrinsic nature of data. Anomalous trends have already been removed in the preprocessing phase. Hence, the smoothing adopted by *f-PCA* had no major effect. Moreover, the discrete temporal grid of acquired kinematic and kinetic data was dense, such that discrete data were comparable with their functional representations.

Therefore, only results obtained from *f-PCA* were reported in the following sections, not to annoy the reader later on, in encountering two times similar results.

5.2 Skill level characterization

Results of only some of the 15 analyzed variables were reported for the athletes' skill characterization: $A_{pr}(t)$, $A_{hs}(t)$, $A_{ks}(t)$, $M_{ks}(t)$, $M_{as}(t)$, $R_{ml}(t)$, $R_v(t)$. The choice of these variables was determined by the discriminant power of their principal component scores in clustering trials of different race walkers.

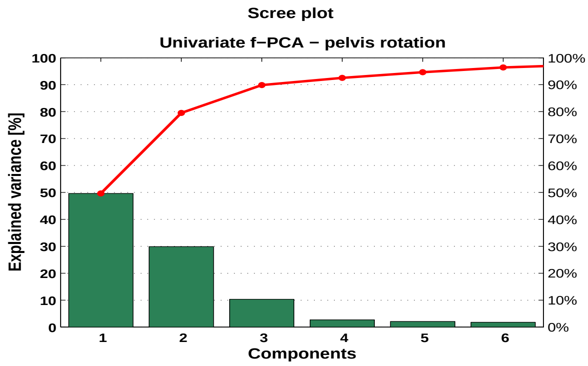


Figure 5.2: Variance explained by the first functional principal components for the pelvis rotation. Each bar represents the variance explained by the corresponding *f-PC*; the line above the bars shows the cumulative percentage.

The scree plot for pelvis rotation was reported in Fig.5.2. The x axis contained the *f-PCs* sorted by decreasing fraction of explained variance. Bars contained the fraction of explained variance. The line above the bars showed the cumulative percentage of explained variability. The first three *f-PCs* explained most (89.85%) of the total variability in the data. Thus they were preserved for further analysis.

First, attention was focused on *f-PC1* and *f-PC2*. In the scatterplot (Fig.5.3(a)), scores for the repeated trials of all the athletes were shown. Different colors were associated to different subjects. The representation of *f-PC2* versus *f-PC1* scores helped to visually evaluate the discriminant power of these components. Scores for *f-PC1* were positive for {s1, s2, s3, s6} and negative for {s4, s5, s7}.

Concerning *f-PC2*, scores did not clearly differentiate athletes. On the contrary, for the same athlete a group of trials scored positively and an other group negatively for *f-PC2*. This was particularly evident for {s5, s6, s7}.

A biomechanical interpretation of the first two *f-PCs* derived by the evaluation of the deviation from the mean (Fig.5.3(b),(c)). The positive scoring on *f-PC1* for {s1, s2, s3, s6} was interpreted by analyzing the red line in Fig.5.3(b). It corresponded to a larger range of movement, with an increased pelvis internal rotation during the swing phase (SP) and increased external rotation in the rear foot support phase (RSP). Conversely the other race walkers, with negative scores on *f-PC1*, were represented by the green line. It showed a restricted internal rotation during SP and an external rotation in RSP.

The increase in pelvic obliquity and in pelvic rotation of the race walkers belonging to group1, demonstrated their cleverness in attenuating the vertical *COM* oscillations. In fact, a pronounced lowering the hip of the moving leg takes place when the supporting leg is fully stretched [98]. Thus, motions of the pelvis in the frontal and transverse planes must tend to minimize the vertical excursions of the *COM*.

Moreover the larger pelvis rotation on the swing side produces a longer stride length for the same amount of hip flexion of the advancing leg and hip extension of the retreating leg [13, 78]. The increase in pelvic obliquity and in pelvic rotation presented by the race walkers of group1 contributed to a faster movement. With the alternating periods of race walking, the pelvis on the front support side rotated to its highest point and the pelvis of the swinging side rotated to its lowest point. The lateral drop of the pelvis away from the stance leg allowed to decrease the horizontal displacement of the *COM*.

To resume, *f-PC1* scores were good discriminant factors among athletes and its interpretation was associated to different ranges of pelvis rotation. Hence, this variable seemed to be decisive in the characterization of athletes of different skill levels.

Looking at Fig.5.3(c), *f-PC2* corresponded a different pelvis angular mean value in the transverse plane throughout the race walking cycle. As previously evidenced, {s5, s6, s7} presented two groups of trials scoring differently for *f-PC2*. Since the two clusters corresponded to trials generated by different limbs (right and left side), these race walkers revealed to have an asymmetrical behavior in pelvis rotation.

Fig.5.4 allowed to interpret the meaning of *f-PC3* and its discriminant potential. The best athlete (s1) was characterized by high positive scores for the third component (Fig.5.4(a)). The score values on this component for the other race walkers might be hardly clustered. *f-PC3* was associated to a time delay in the passage from external to internal pelvis rotation (Fig.5.4(c)). Since this characteristic was strictly related to the knee extension, it underlined the ability of s1 in straightening the leg for a long time, thus getting an extra thrust forward.

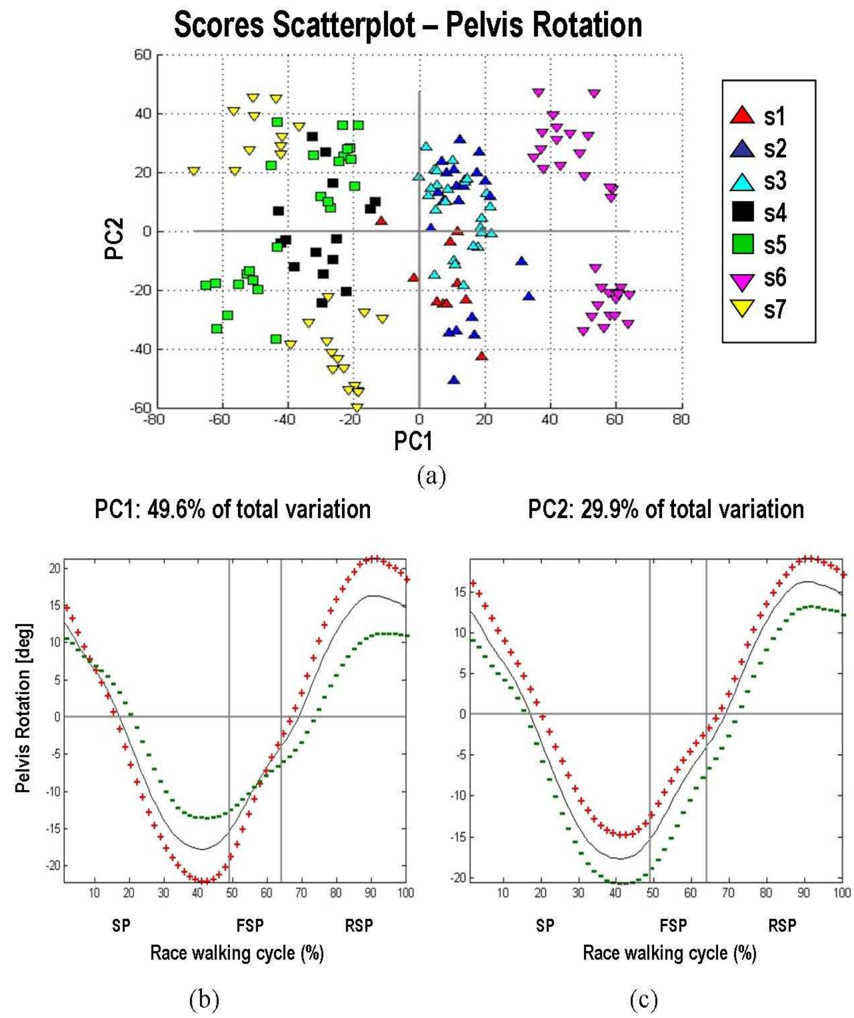


Figure 5.3: Characterization of pelvis rotation: (a) scores scatterplot of *f-PC2* vs *f-PC1*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c). The horizontal line is the zero axis; the vertical lines divides race walking phases.

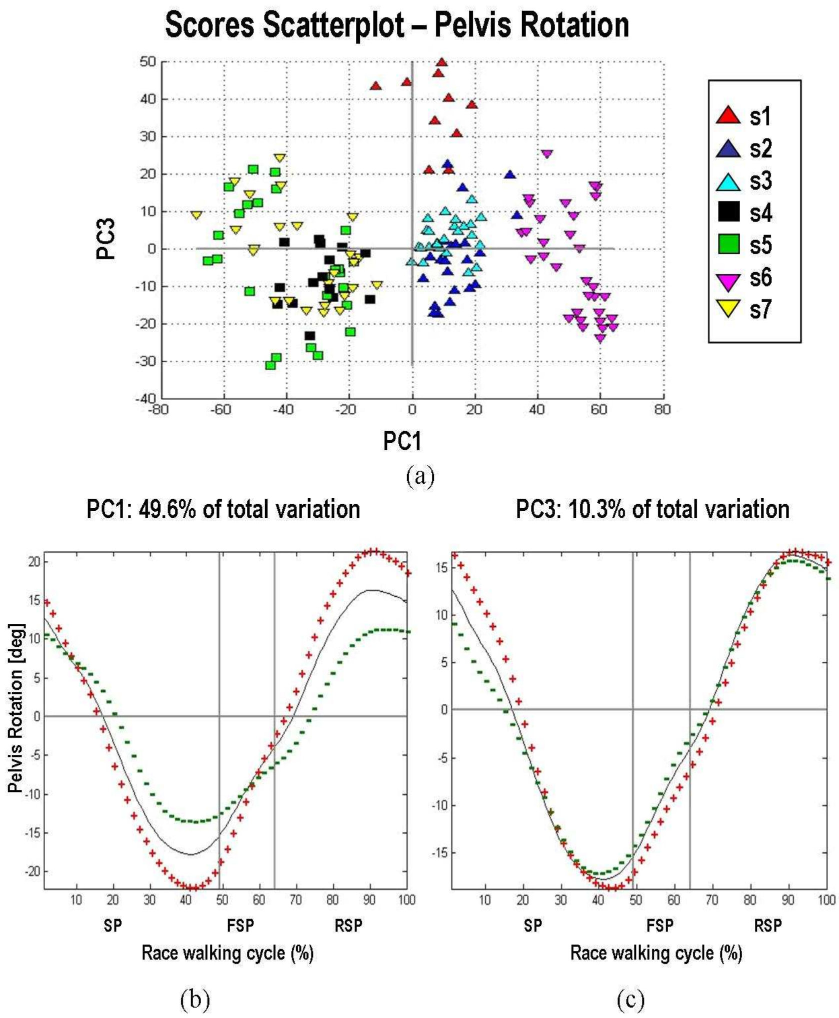


Figure 5.4: Characterization of pelvis rotation: (a) scores scatterplot of *f-PC3* vs *f-PC1*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC3* (c). The horizontal line is the zero axis; the vertical lines divides race walking phases.

The hip joint in the sagittal plane presented interesting results. The scree plot was reported in Fig.5.5. The first two components accounted for 83.76% of data variance. The barplot showed how the first component dominated (74%). It underlined the great importance of *f-PC1* in describing the main behavior in the athletes' hip flexion-extension.

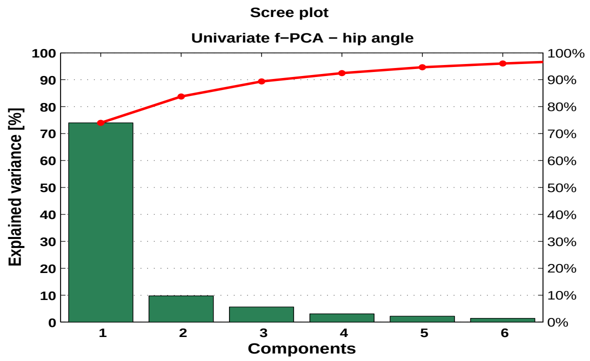


Figure 5.5: Variance explained by the first functional principal components for the hip angle in the sagittal plane. Each bar represents the variance explained by the corresponding *f-PC*; the line above the bars shows the cumulative percentage.

Moreover the corresponding scatterplot (Fig.5.6(a)) showed an high inter-subject discriminant power both for *f-PC1* and *f-PC2*. Concerning the x axis (*f-PC1*), the best athlete (s1) had high positive scores, while one of the least performing ones (s6) showed high negative scores. A biomechanical interpretation of the first functional principal component derived from Fig.5.6(b). *f-PC1* revealed to be related to a postural behavior: athletes scoring positively (red curve) maintained their hip more flexed than the mean curve throughout the movement. On the contrary s6, scoring negatively, extended the hip during the whole race walking cycle.

Considerable hip flexion throughout the race walking cycle indicates a forward shifting of the *COM*, crucial factor for forward propulsion [13, 78]. Moreover it allows the race walker to walk with a fluent style [115]. On the contrary, bending from the waist causes the hips to move backwards, thus reducing the ability to extend the hip and accelerate the stride. Hence, one of the factors that more differentiate s1 and s6 in performing level might be ascribed to a different behavior in hip flexion and extension. The *f-PC2* was a discriminant factor for the other athletes {s2, s4, s5}: positive scores for s2, while negative for s4 and s5. This functional component was related to a technical factor (Fig.5.6(c)). The athletes of group2 {s4, s5}, characterized by negative *f-PC2* scores, adopted the strategy to bend the hip less than the average angle during the swing phase (SF), and extending it less during the rear leg support phase (RSF).

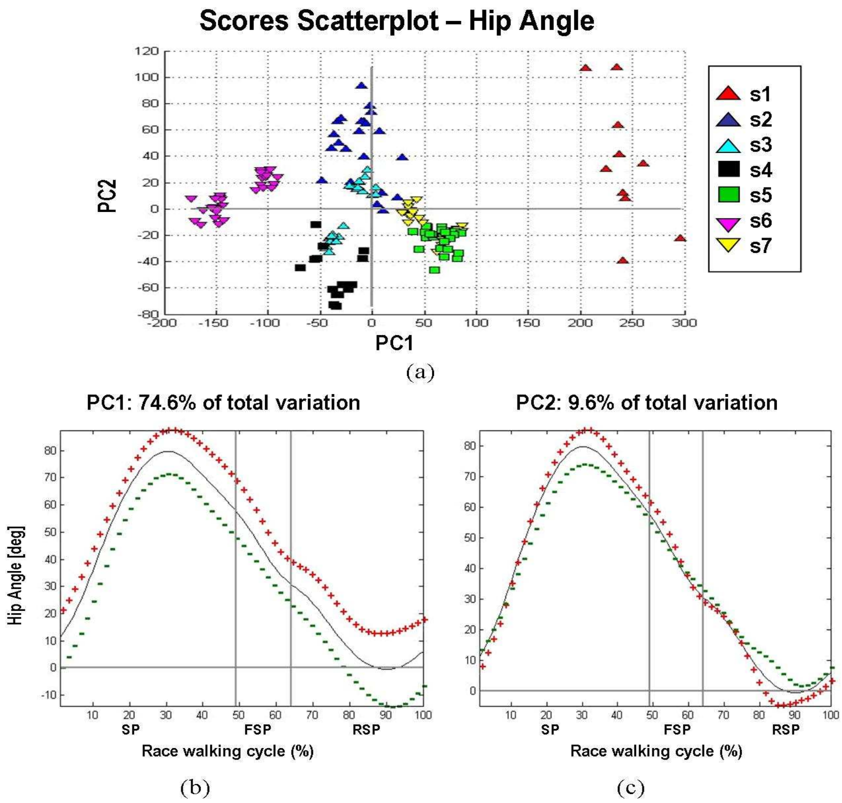


Figure 5.6: Characterization of hip joint in the sagittal plane: (a) scores scatterplot of *f-PC2* vs *f-PC1*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c). The horizontal line is the zero axis; the vertical lines divides race walking phases.

This is a counterproductive behavior, since the hip extension in late swing helps the walker to increase the joint angle amplitude and to gain momentum to pull the body forward [78].

As reported by Cairns [13], race walkers display a huge amount of hip flexion during the swing phase of race walking, which becomes significantly greater with the increase of velocity.

Particular attention must be dedicated to the knee joint angle in the sagittal plane. The requirement of the International Federation rules to keep the supporting leg in a straight position, make the knee angle to differ significantly respect to its behavior in normal gait analysis [13, 78, 90, 91, 115].

The most relevant discrepancies appear during load acceptance and midstance: the normal knee flexion that occurs in the first phase of stance disappears and an hyperextension angle is maintained from about 25% to 75% of contact time [91].

The first three functional principal components accounted for most of the variance in the data (93.7%), as reported in Fig.5.7. Hence further investigations were restricted to these *f-PCs*.

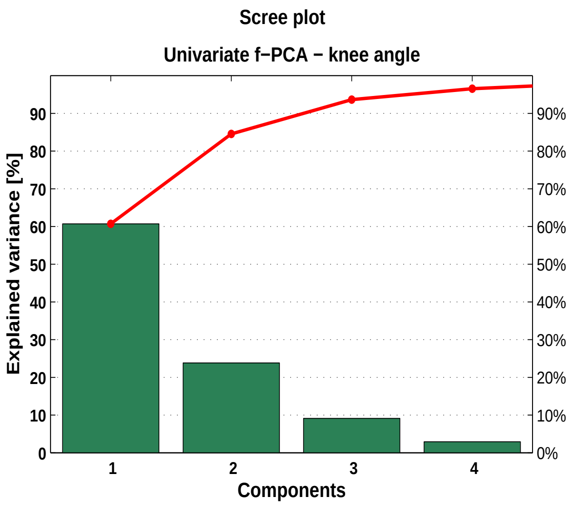


Figure 5.7: Variance explained by the first functional principal components for the knee angle in the sagittal plane. Each bar represents the variance explained by the corresponding *f-PC*; the line above the bars shows the cumulative percentage.

The first functional component (Fig.5.8(b)) represented athletes with different mean values for the flexion-extension of the knee throughout the race walking cycle. Therefore it was interpreted as a postural difference among race walkers. Especially s6 differentiated from the others: with high negative scores for *f-PC1*, this subject exhibited a more extended knee throughout the movement.

The best athletes {s1, s2} showed (Fig.5.8(a)) opposite behaviors for the *f-PC2*: s1 scoring positively, while s2 negatively. This result seemed to contrast with the hypothesis of similar technical characteristics for race walkers belonging to analog skill levels (group1).

A possible explanation derived by the analysis of *f-PC2* and *f-PC3*. They allowed to reveal how two subjects of the same skill level adopted completely different motor strategies. Fig.5.9(b) demonstrated how s1, with positive *f-PC2* scores, maintained an extended knee (knee joint angle near to zero value) during the stance phase, while s2, with negative scores, tended to hyperextend the knee (negative joint angle) respect to the average curve during the stance phase.

A better explanation of this difference derived from the interpretation of *f-PC3* (Fig.5.9(c)). s1 showed a smoother trend for knee flexion-extension (red line), with a smaller range of movement. This athlete extended the knee just before the heel strike and got it hyperextended only in the late stance phase. The faster preparation of the knee derived from a lower flexion during the SP. On the contrary, s2 took longer to get to knee full extension.

For both the athletes, leg was straightened for a long time after the passage through the vertical position, thus getting an extra thrust forward by pushing off the rear foot [13, 78, 91]. Fig.5.9(a) emphasized the strong relation between *f-PC2* and *f-PC3*: most of the score values arranged around a diagonal line through the plot. In fact, while *f-PC2* was related to a general technical characteristic of the SF, *f-PC3* gave a more subtle description of the fractional parts of the same phase.

The process of hyperextension for athletes {s2, s4, s6, s7} might put stress on the posterior structures of the knee joint [78, 91]. Over a long period of time, this stress may be injurious to the ligaments. To provide dynamic ligamentous support against excessive hyperextension, strengthening of the knee flexor muscle is usually recommended. Moreover, concerning *f-PC2*, for {s5, s6, s7}, two separate clusters of points could be clearly recognized, corresponding to trials evaluated for the two lower limbs (right and left side). This implies that these race walkers had an asymmetrical movement for the dominant and non dominant legs. To resume, less performing race walkers (group3) showed knee hyperextension and asymmetry between right and left limb, thus revealing a possible risk of injury.

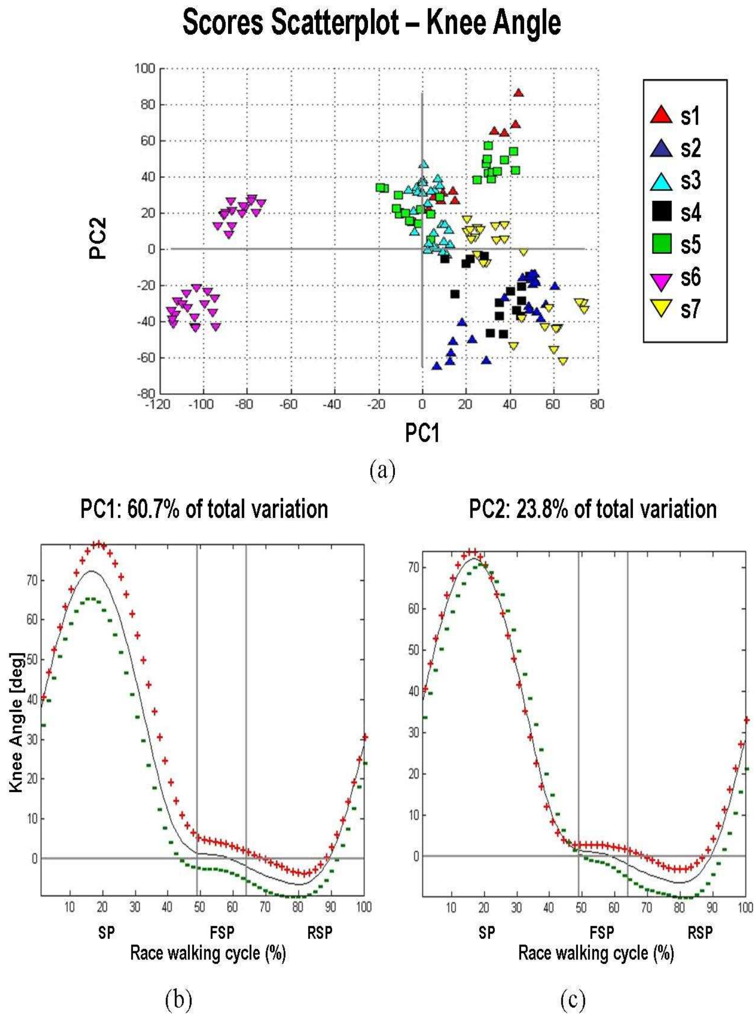


Figure 5.8: Characterization of knee joint angle in the sagittal plane: (a) scores scatterplot of *f-PC2* vs *f-PC1*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c). The horizontal line is the zero axis; the vertical lines divides race walking phases.

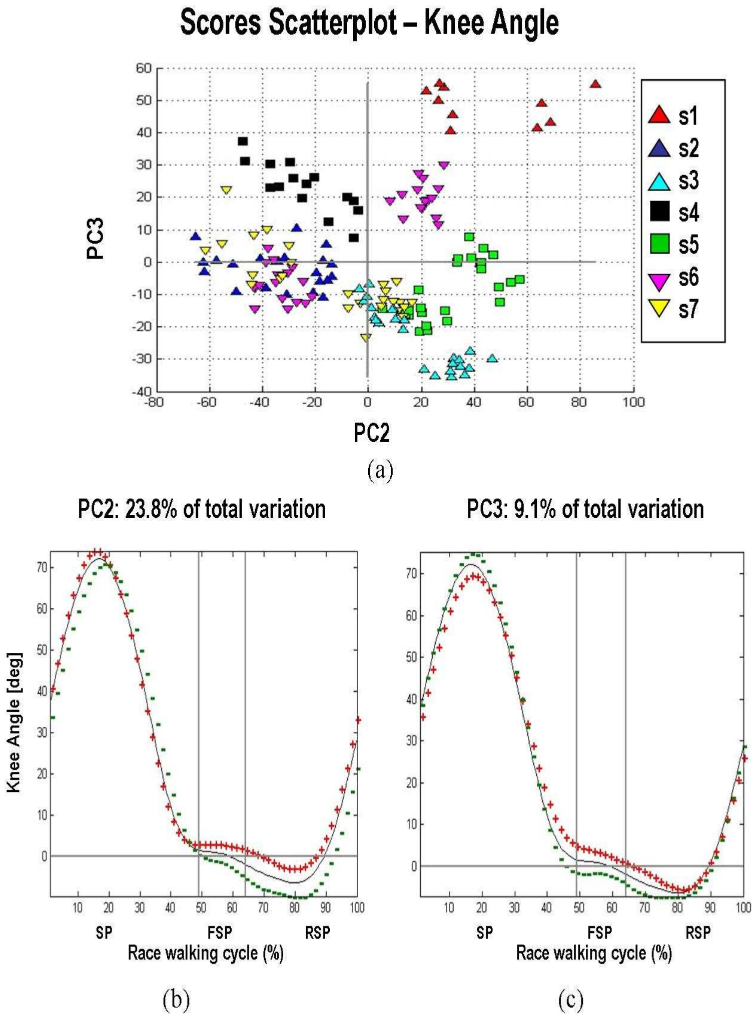


Figure 5.9: Characterization of knee joint angle in the sagittal plane: (a) scores scatterplot of *f-PC3* vs *f-PC1*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC3* (c). The horizontal line is the zero axis; the vertical lines divides race walking phases.

Also the knee joint moment revealed to be a good discriminant factor in characterizing athletes according to their skill levels. The first four functional components were investigated; they accounted for 90.2% of data variance (Fig.5.10).

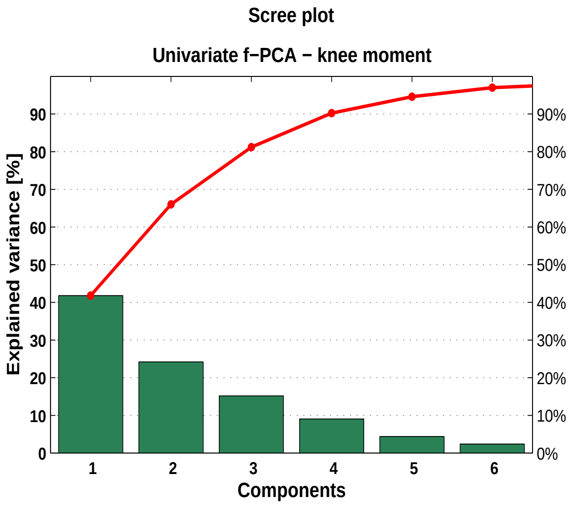


Figure 5.10: Variance explained by the first functional principal components for the knee moment in the sagittal plane. Each bar represents the variance explained by the corresponding *f-PC*; the line above the bars shows the cumulative percentage.

The group1 scored positively for the *f-PC1* (Fig.5.11(a)), thus revealing an higher knee extension moment than the mean curve during the stance phase (Fig. 5.11(b)). It can be seen as the effort exerted by knee extensor muscles to keep that joint straight, just like rules impose.

A flexion moment occurred in the mean trend at the time corresponding closely to the heel strike, to compensate the external hyperextension moment, which is the moment present at knee angle of hyperextension. This flexion moment was higher in athletes scoring positively for the *f-PC2* (Fig.5.11(c)), especially for {s3, s5}.

The knee flexion moment is interpreted by some authors [13, 78] as the outcome of passive structures (posterior capsule and ligaments) rather than active muscular forces. Therefore the knee joint angle and moment should be monitored to prevent injuries. Athlete s4, scoring negatively for *f-PC2*, showed a lower knee extension moment than the mean curve at the toe-off.

The extension forces at the knee are of great importance in achieving the push-off in race walking [13].

Fig.5.12(a) showed how *f-PC3* and *f-PC4* contributed to athletes' discrimination. s4 and s7 scored negatively on *f-PC3*, thus revealing (Fig.5.12(b)) a tendency to postpone the initial knee extension moment and to maintain the delay during the stance phase. A subtle difference between these athletes was pointed out by *f-PC4* (Fig.5.12(c)). This functional component revealed the knee behaving in different ways for s4, scoring positively, and s7, scoring negatively. Mainly, the less performing race walker appeared to have difficulties in maintaining a correct knee flexion moment just after the passage through the vertical position. Hence, the green line in Fig.5.12(c) (negative scores) had low smoothness.

The asymmetries between right and left limb, evidenced in the knee angle analysis, were evident also in the knee flexion moment. Separated scores clusters could be seen for s6 (Fig.5.11(a)) and s7 (Fig.5.12(a)).

The knee joint angle is directly related to the ankle movement. In race walking, ankle dorsiflexion occurs just prior to heel strike. During the double support phase, the *COM* is at its lowest height and the ankle helps in raising it. Hence the so called "functional lengthening" [78] occurs. It consists in an increased stride length: at the rear limb the ankle plantarflexes; at the forward limb the ankle dorsiflexes and the knee extends [13, 78, 90, 115].

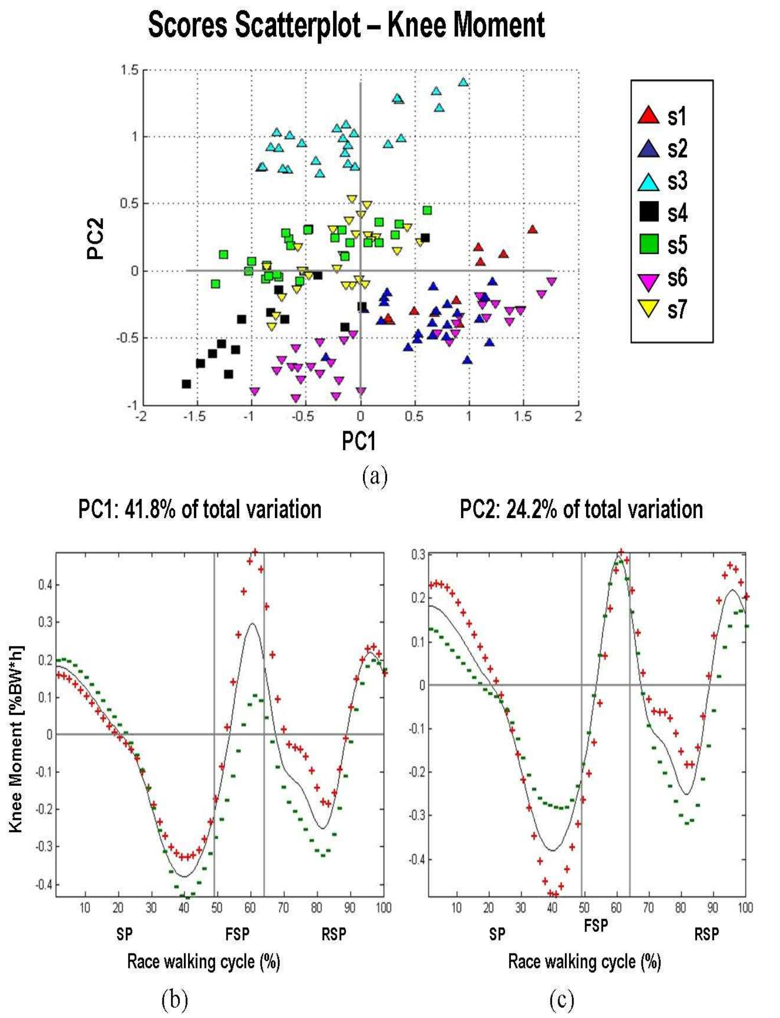


Figure 5.11: Characterization of knee joint moment in the sagittal plane: (a) scores scatterplot of *f-PC2* vs *f-PC1*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c). The horizontal line is the zero axis; the vertical lines divides race walking phases.

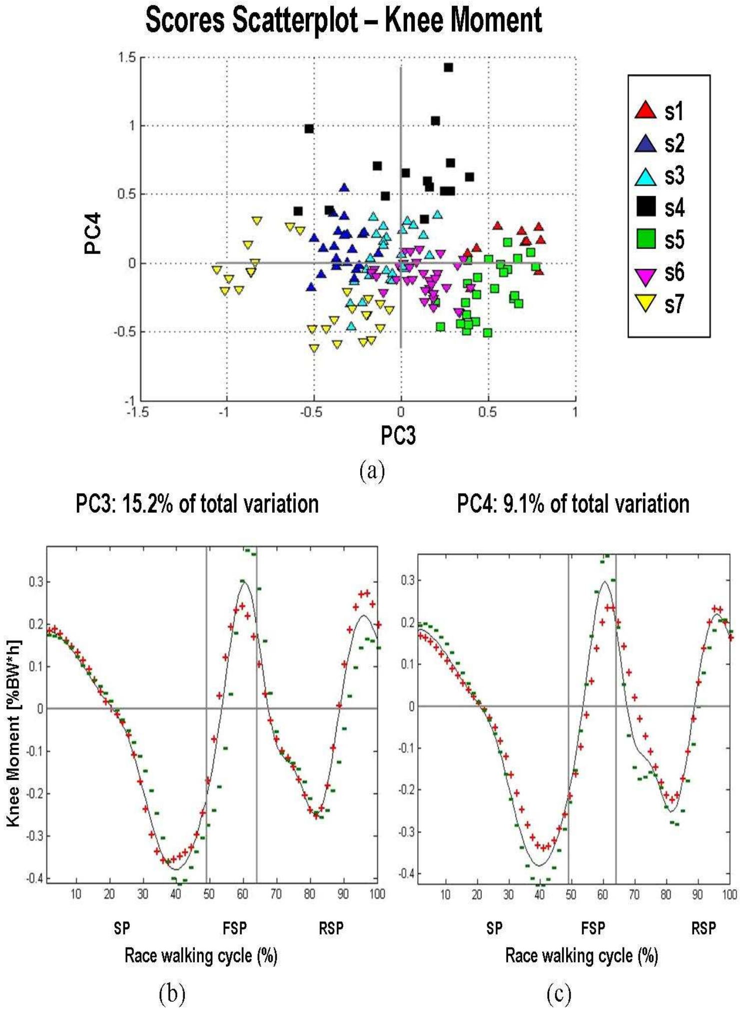


Figure 5.12: Characterization of knee joint moment in the sagittal plane: (a) scores scatterplot of *f-PC4* vs *f-PC3*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC3* (b), *f-PC4* (c). The horizontal line is the zero axis; the vertical lines divides race walking phases.

The first two *f-PCs* for the ankle moment in the sagittal plane explained about 86.7% of the total variance (Fig.5.13).

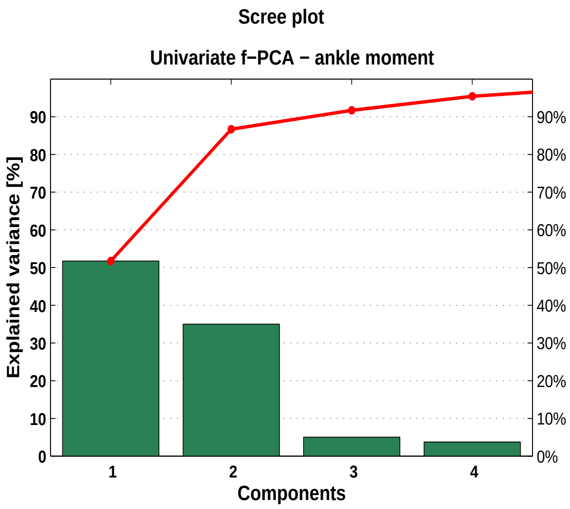


Figure 5.13: Variance explained by the first functional principal components for the ankle moment in the sagittal plane. Each bar represents the variance explained by the corresponding *f-PC*; the line above the bars shows the cumulative percentage.

Fig.5.14(a) showed how s4 had high positive scores for *f-PC2*. The biomechanical interpretation (Fig.5.14(c)) helped in explaining this behavior as a lower dorsiflexion moment at the ankle during the FSP and an higher plantarflexion moment during the RSP.

Race walkers {s2, s3, s4} scored negatively for *f-PC1* (Fig.5.14(a)), thus revealing a tendency to postpone the ankle dorsiflexion moment after the heel strike and to maintain a delay in the plantarflexion during the stance phase (Fig.5.14(b)).

The dorsiflexion moment exerted during the early stance phase results from eccentric muscle contraction used to decelerate forward progression of the tibia [78]. The delay in dorsiflexion moment might imply a greater amount of ankle plantarflexion moment in the last stance phase, to prepare the foot for the toe-off.

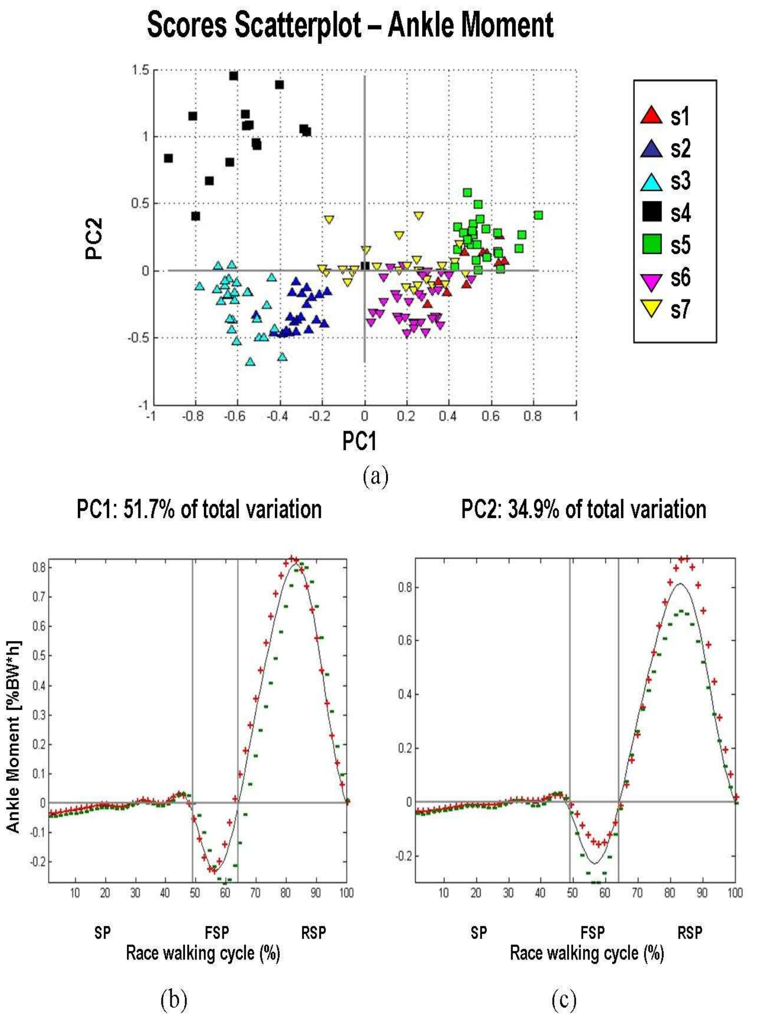


Figure 5.14: Characterization of ankle joint moment in the sagittal plane: (a) scores scatterplot of *f-PC2* vs *f-PC1*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c). The horizontal line is the zero axis; the vertical lines divides race walking phases.

The plantarflexion moment is thus exerted for a shorter period of time, thus assisting the athlete to maintain a straight knee. In fact both ankle and knee joint angle influence the knee joint flexion moment produced by the gastrocnemius. A plantarflexed ankle, associated to a straight position of the knee, facilitates the intervention of the gastrocnemius [91]. Thus good race walkers should exploit as soon as possible the contribution of this muscle.

In race walking a medial GRF greater than in normal walking is achieved early and late in the stance phase to contrast the lateral shift of the *COM*, due to the pelvis lateral drop and hip adduction [13, 91]. Fig.5.15 demonstrated how the first three components were able to explain most (93.9%) of the variance.

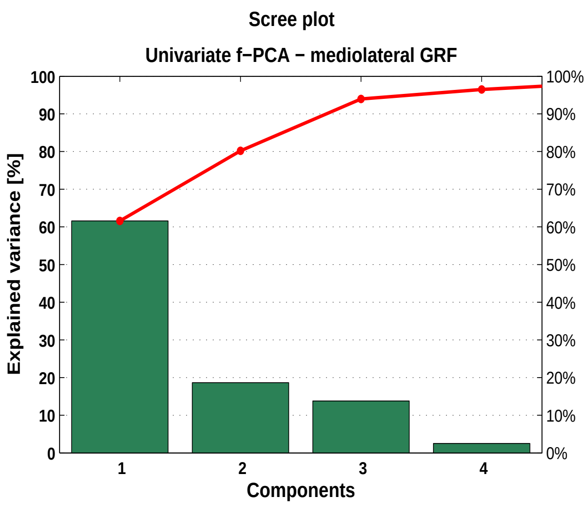


Figure 5.15: Variance explained by the first functional principal components for the mediolateral GRF. Each bar represents the variance explained by the corresponding *f-PC*; the line above the bars shows the cumulative percentage.

Race walkers ((Fig.5.16(b) black mean curve) involved a quite sensitive oscillation in the mediolateral direction, even if the sum of positive and negative areas resulted nearly close to zero [91]. That indicated that the subjects with a mediolateral force curve resembling the mean trend had a correct technique, with straight progression.

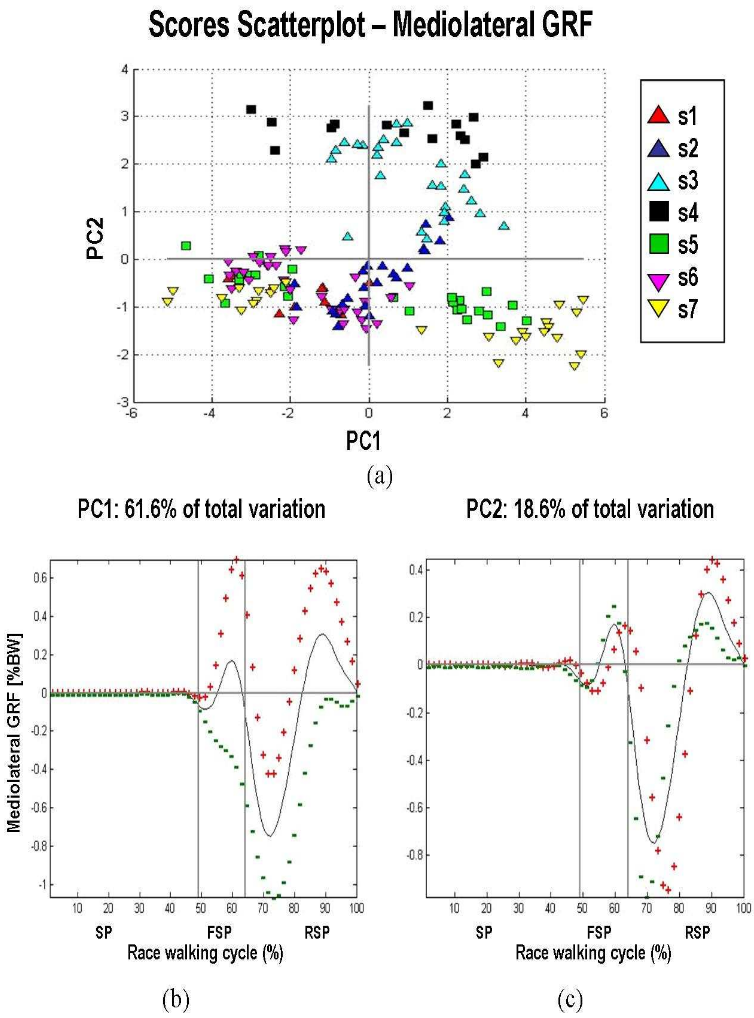


Figure 5.16: Characterization of mediolateral GRF: (a) scores scatterplot of *f-PC2* vs *f-PC1*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c). The horizontal line is the zero axis; the vertical lines divides race walking phases.

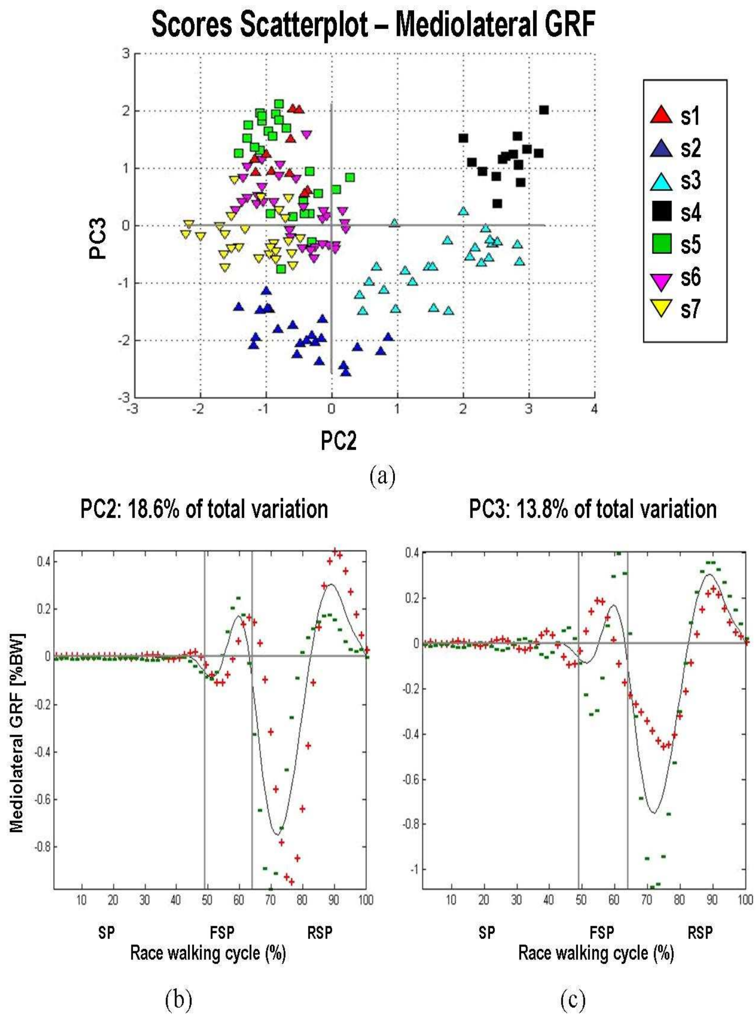


Figure 5.17: Characterization of mediolateral GRF: (a) scores scatterplot of *f-PC3* vs *f-PC2*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC2* (b), *f-PC3* (c). The horizontal line is the zero axis; the vertical lines divides race walking phases.

Athletes s1 and s2 presented a very small range of score values both for the first and the second *f-PCs*, thus highlighting their ability in performing the movement with great repeatability. A shift in time compared to the mean trend is well recognized when comparing athletes scoring positively (delay) with the ones with negative scores (anticipation) for *f-PC2* (Fig.5.16(c)).

Athletes s3 and s4, with high positive scores on *f-PC2*, showed how the negative and the second positive peak were higher than the mean trend in the mediolateral GRF. The second peak was higher, to increase the deceleration of the lateral shift of the pelvis for preparing the next stance phase [13, 78].

Also *f-PC3* demonstrated a great discriminant power (Fig.5.17(a)). Actually, the scatterplot of *f-PC3* vs *f-PC2* scores clearly identified clusters of trials belonging to different subjects. {s4, s5, s6}, scoring positively on *f-PC3*, showed an advanced first medial peak and a very reduced lateral peak. {s2, s3}, with negative *f-PC3* scores, revealed the first positive and the negative peak to be higher than the mean trend.

The vertical GRF increases significantly across gait conditions from walk to race walk to run [13, 78, 81, 90, 91]. Most (93.5%) of data variance was explained by the first three functional components (Fig.5.18).

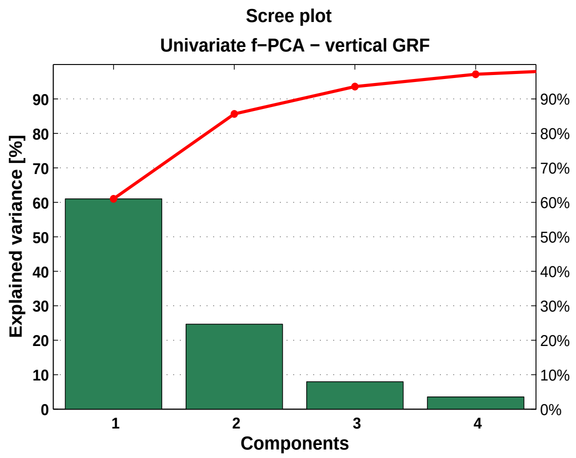


Figure 5.18: Variance explained by the first functional principal components for the vertical GRF. Each bar represents the variance explained by the corresponding *f-PC*; the line above the bars shows the cumulative percentage.

Athletes {s1, s5}, with high negative scores for *f-PC1* (Fig.5.19(a)), were characterized by force peak higher than the mean curve during the FSP and a lower one in the RSP (Fig.5.19(b)).

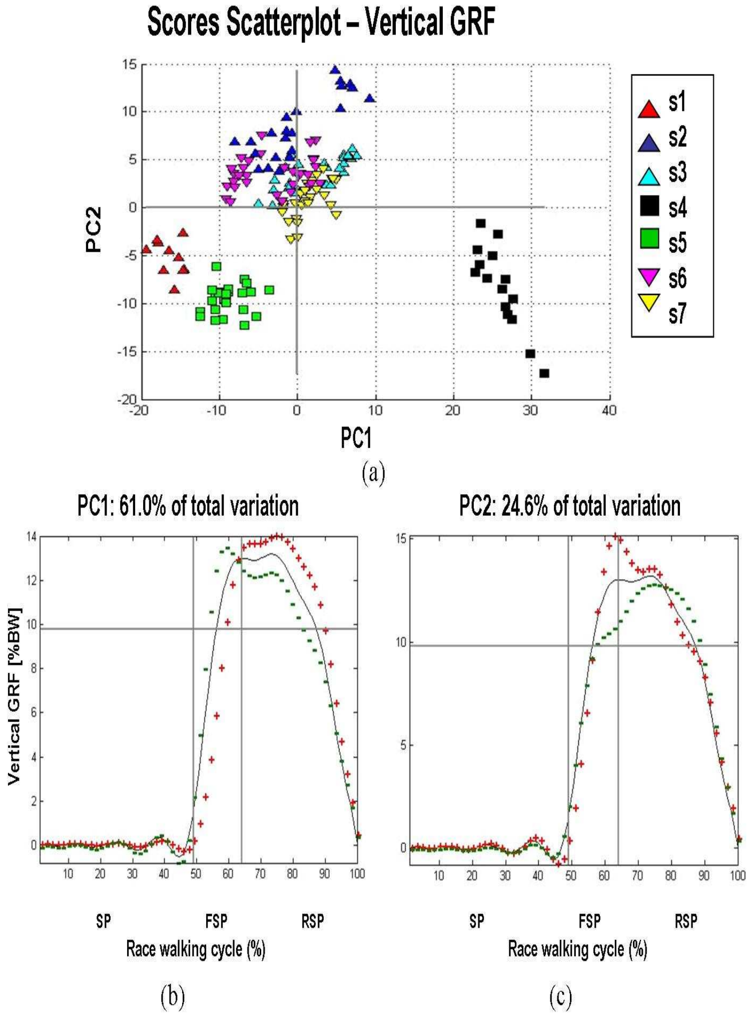


Figure 5.19: Characterization of vertical GRF: (a) scores scatterplot of *f-PC2* vs *f-PC1*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c). The horizontal line is the 9.81 axis; the vertical lines divides race walking phases.

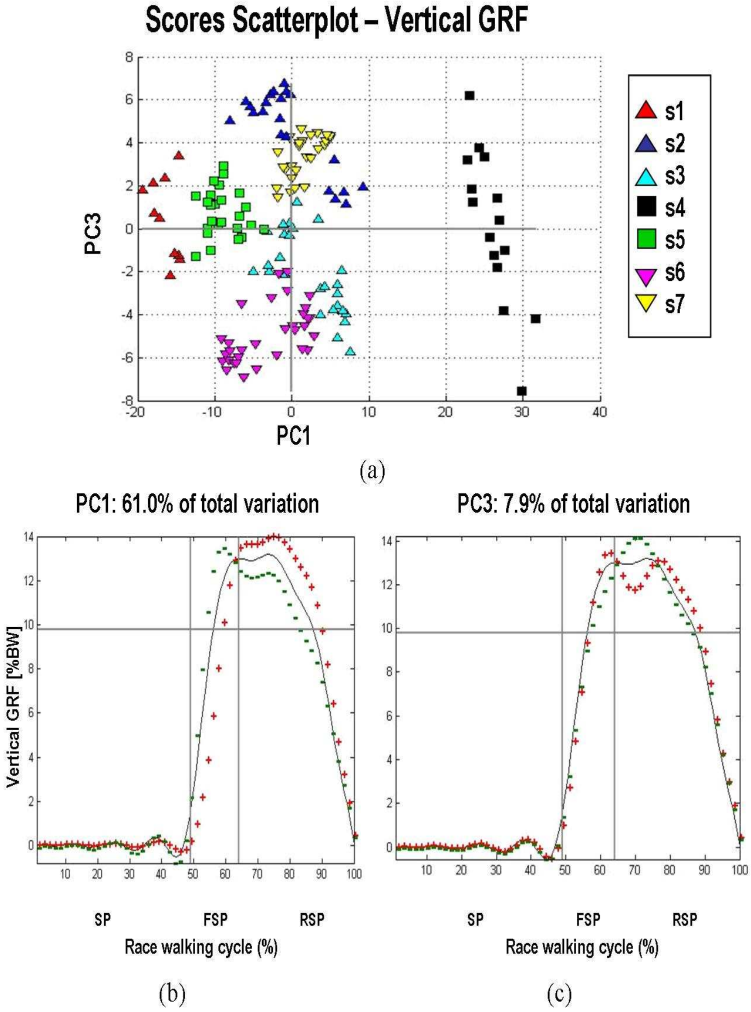


Figure 5.20: Characterization of vertical GRF: (a) scores scatterplot of *f-PC3* vs *f-PC1*; (b,c) the mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC3* (c). The horizontal line is the 9.81 axis; the vertical lines divides race walking phases.

The first peak might be due to an higher race walking velocity, involving a greater impact with the ground.

In the last stance, the vertical GRF became inferior respect to the mean trend probably to decrease the upward thrust of the *COM*. On the contrary, s4, with very high positive *f-PC1* scores, showed a more smooth force trend, where a central plateau replaced the typical walking curve with two peaks and a midstance valley. This proved the ability of this athlete in reducing the vertical acceleration of the *COM*, thus achieving a less expensive action.

The other race walkers {s2, s3, s6, s7}, scoring positively on the *f-PC2*, showed a peak vertical GRF in the passage through the vertical position, i.e from FSP to RSP (Fig.5.19(c)).

A clear-cut division among these athletes was extracted by the analysis of *f-PC3* (Fig.5.20(a)). s2 and s7 scored positively, while s3 and s6 scored negatively on *f-PC3*. Looking at Fig.5.20(c), the first two race walkers presented a vertical force pattern similar to the normal walking one, while the other two resembled the running trend.

s4 had a very large range of *f-PC3* scores, thus revealing a low reproducibility in performing the race walking movement.

5.3 Coordinative characterization

Up to here, the attention was focused on single variable description, with the intent to gain more insight onto the characterization of individual athlete peculiarities. A further step consists in the evaluation of how lower limb joints interact to perform an efficient movement. Principal component analysis revealed good potentialities also for the coordinative characterization of race walkers. Two-stage and bivariate functional PCA were adopted. To avoid misinterpretation, the principal components measured in the *2-PCA* were named *2-PCs*, i.e. the first component was *2-PC1* and the second one *2-PC2*.

According to dynamic system theory (DST), human limbs can be seen as a system of coupled pendulums that oscillate around joints [12, 17, 110]. Changes in the mutual relations between body segments or adjacent joints may give important indications about the inherent coordinative factors of the locomotor system.

The choice to insert the angular velocities in the analysis of coordination derived from their undoubted importance in the coordinative characterization of a movement. Moreover, entering these variables in the PCA analysis, results resembled those derived from continuous phasing relationships (CRP), thus enabling the comparison of results. CRP has been largely adopted in literature [114, 68, 67, 89] for the analysis of different elements that participate to the movement. It allows to synthesize information, condensing two variables within one and investigating synergies between interacting joints. In fact, CRP consists in phase plots of the angle on the horizontal axis, with its first derivative (the angular velocity) on the vertical axis.

To evaluate the degree of interaction of different variables describing an athlete movement, *2-PCA* was applied to fifteen variables: 3 pelvic rotations, 3 joint angles and 3 moments in the sagittal plane, 3 GRF components and 3 joint angular velocities in the sagittal plane (hip, knee, ankle).

The purpose of *2-PCA* application was the detection of the first two main "motion features" that characterize the relative motion of body parts in race walking. The loadings of the first two components on each of the 15 kinematic and kinetic variables were reported in Fig.5.21. They represented the relative contribution of each variable to the first stage *2-PC1* and *2-PC2*.

$A_{pr}(t)$, $A_{hs}(t)$, $M_{hs}(t)$ and $V_{ks}(t)$ contributed to the first feature (Fig. 5.21(a)). The signs of the loadings helped to understand the mutual relations among variables. The loadings of *2-PC1* on $A_{pr}(t)$ and $A_{hs}(t)$ were approximately equal and opposite. This indicated that when the pelvis was intra-rotated, the hip was flexed, when the pelvis was extra-rotated the hip was extended. $V_{ks}(t)$ had opposite sign respect to $A_{hs}(t)$. This means that

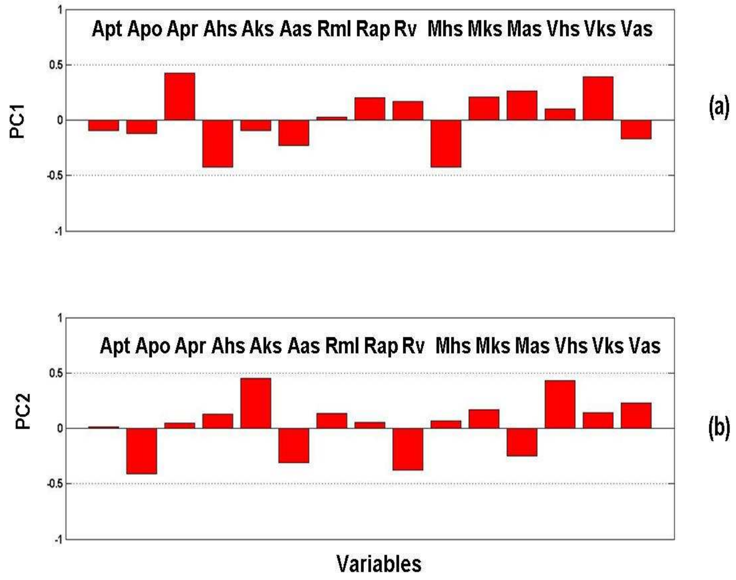


Figure 5.21: Loadings of the first two *2-PCs* on each of the fifteen kinematic and kinetic variables.

when hip angle was extended, knee angle velocity was positive (flexion); similarly when hip angle was flexed, knee angle velocity was negative (extension). Thus, the first motion feature detected the strong importance and interaction of only few factors in race walking description: hip and knee movement in the sagittal plane and pelvis rotation.

More variables contributed to the second component (Fig. 5.21(b)): $A_{po}(t)$, $A_{ks}(t)$, $A_{as}(t)$, $R_v(t)$, $M_{as}(t)$ and $V_{hs}(t)$. The loadings revealed one more time the strict relation between knee angle and hip angular velocity. Moreover, the pelvis obliquity and the ankle flexion showed a great importance in the definition of the second motion feature. Signs gave indications on the synergies of joints: when the knee was extended, the ankle was dorsiflexed, while the hip was high in the sagittal plane.

Interestingly, *f-PCA* and *2-PCA* analysis led to analog results. In fact, the fundamental variables for the description of race walking motion features, were the most discriminant ones in the skill level characterization (Section 5.2).

2-PCA application was then restricted only to some kinematic variables: hip, knee, ankle joint angles and angular velocities in the sagittal plane. This choice derived from the necessity to simplify the interpretability of results. Graphics involving only angles and angular velocities should better resemble CRP representations. This might facilitate the comprehension of the athletes' coordination strategies.

With this new set of variables, the first three components allowed to account for about 90% of the total data variance for the first stage of *2-PCA*. In particular, the first two components revealed to be the most representative ones (Fig.5.22).

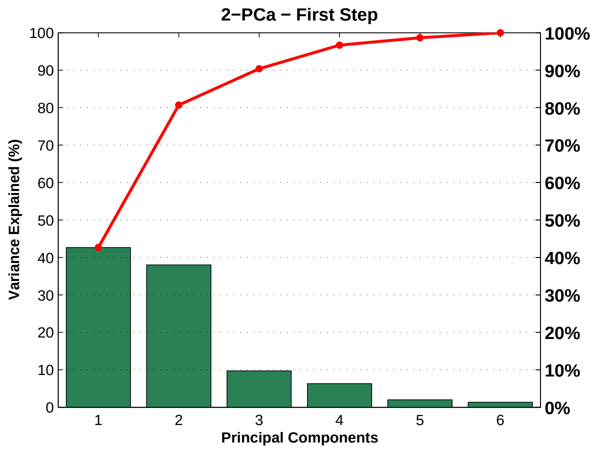


Figure 5.22: Scree plot showing the percentage of variation in the data explained by the principal components.

A biomechanical interpretation of these *2-PCs* derived from the inspection of the contribution on each of the six kinematic variables (Fig.5.23). Hip and knee revealed their great importance in the definition of race walking movement: $A_{ks}(t)$, $A_{as}(t)$ and $V_{hs}(t)$ were at the base of the first motion feature for *2-PC1* (Fig.5.23(a)); $A_{hs}(t)$, $V_{ks}(t)$ and $V_{as}(t)$ contributed to the second motion feature for *2-PC2* (Fig.5.23(b)).

The projection space of data on these components (*2-PC1*, *2-PC2*) was similar to a "O"-shaped manifold, as reported in Fig.5.24. The different race walking cycle phases were defined by different colors: swing phase [AB] in red, frontal leg support phase [BC] in green, rear leg support phase [CD] in cyan. The "O"-shaped manifold was traversed clockwise, once during every gait cycle.

The first stage *2-PC2* seemed to be related to differences between the swing phase and the stance phase. In fact, positive data projections on this

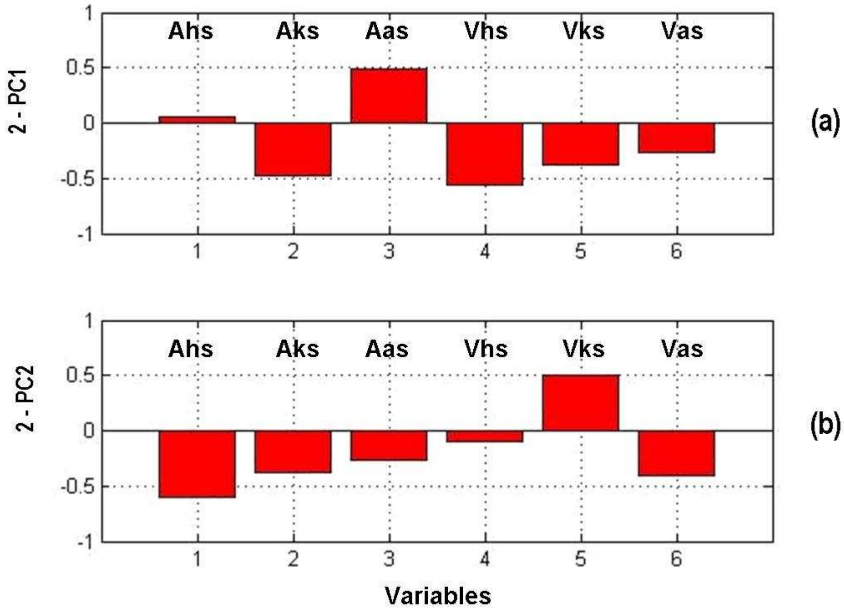


Figure 5.23: Loadings of the first two *2-PCs* on each of the six kinematic variables.

component (right side of Fig.5.24) concerned the stance phase and the first part of swing; negative projections (left side of the figure) were related only to the swing phase. This *2-PCA* component was, as previously described, strictly related to the hip angle and to the knee and ankle angular velocities in the sagittal plane (Fig.5.23(b)). This implies that the second motion feature of race walking corresponded to the behaviour of the hip in the different phases of race walking: it was flexed throughout the stance phase and the initial swing and extended elsewhere. The hip movement was obviously related to knee and ankle angular velocities.

Concerning the first stage *2-PC1*, positive projections (upper part of Fig.5.24) represented the race walking phase that goes from the terminal swing to the early RSP. As previously discovered, variables involved in the first motion feature were $A_{ks}(t)$, $A_{as}(t)$ and $V_{hs}(t)$ (Fig.5.23(a)). Thus *2-PC1* corresponded to the interaction among these variables when approaching the ground and passing through the vertical position. When the knee extended, the ankle dorsiflexed and a rapid hip extension was carried out. This behavior identified a mean of functional lengthening, resulting in forward projection

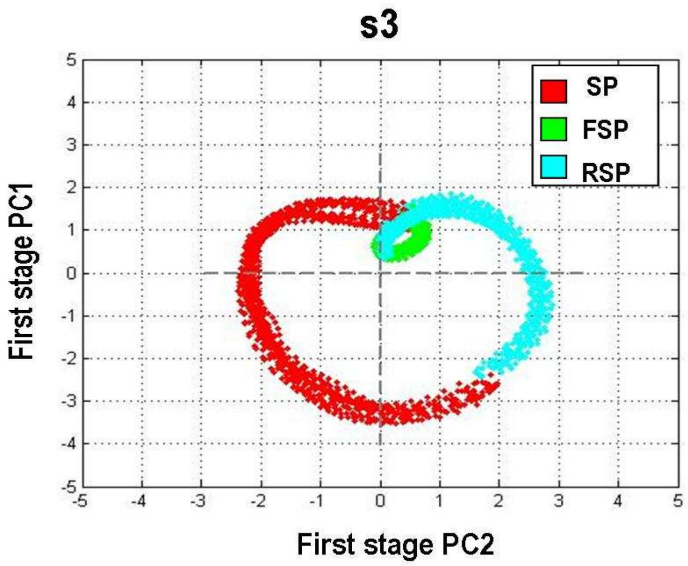


Figure 5.24: Example of manifold in {*2-PC1*, *2-PC2*} space after first stage of *2-PCA* for s3.

of the calcaneus and increased stride length [78, 13]. Hence, the first stage *2-PC1* was a good index of athletes' skill level, because it focused the attention of the most important phase for race walkers.

Fig.5.25 compared the "O"-shaped manifolds of athletes s3 and s6. Since athletes were very high level race walkers, variations in the shape were quite subtle. Little differences might be inferred from the two "O"-shaped manifolds and from how the manifolds were traversed during a cycle. The best athlete (s3-group1) showed great repeatability of movement, especially during the stance phase. In fact, manifold corresponding to different race walking cycles were closed each other. Otherwise, s6 (group3) had a rather scattered trend among different trials, thus revealing a difficulty in performing the race walking movement always in the same way. Two different trends were clearly visible in the manifolds of this athlete, during the swing phase. They corresponded to trials for right and left limbs, thus confirming the asymmetric movement of this race walkers, already shown in Section 5.2.

A clear-cut difference between these two athletes were showed by the first stage *2-PC1*. Positive projections on this component had a much more smooth trend for the best athlete than for the less performing one. This result confirmed the great importance of this phase of race walking cycle in determining the level of performance of the athlete.

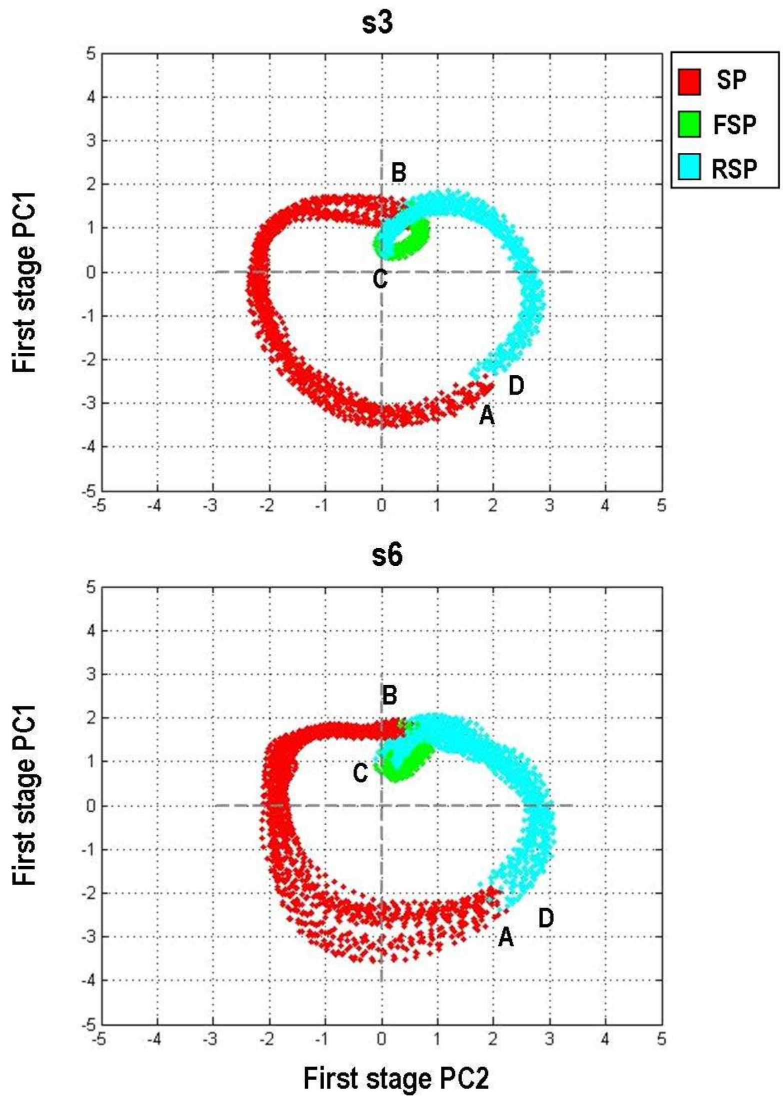


Figure 5.25: Example of manifolds in (*2-PC1*, *2-PC2*) space after first stage of *2-PCA*: s3 and s6.

For the second-stage *2-PCA*, the first two components derived from the first stage (*2-PC1*, *2-PC2*) were concatenated and inserted in a second PCA. This analysis led to the recognition of the seven athletes, as shown in Fig.5.26. All the athletes were recognizable in separated clusters. In particular, s1 and s6 showed very peculiar behaviors, in agreement with the results (Section 5.2). In fact, the univariate *f-PCA* had shown that these subjects differentiated most in the hip flexion-extension (Fig.5.6), which thus played a central role in the whole movement determination. From the two-stage *2-PCA*, it was difficult to give a biomechanical interpretation of differences in clusters. In fact, the application of PCA twice and the concatenation of two components of the first stage, involved the loss of a full understanding of the relationships among kinematic and kinetic variables in determining the second stage components. Thus, even if the scatterplot showed clear-cut divisions among clsters, it was hard to comprehend where the athletes differed the most.

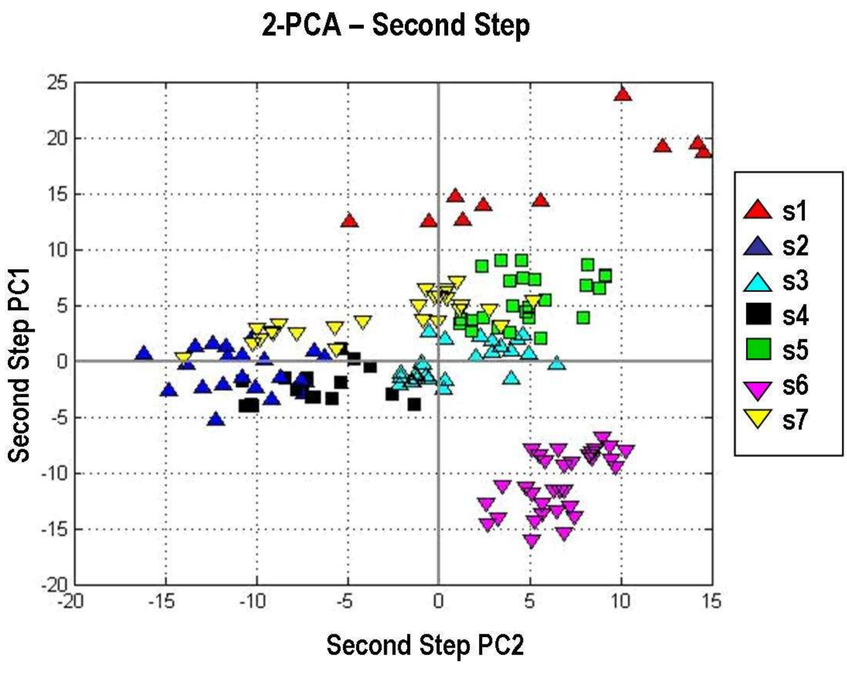


Figure 5.26: Discrimination of athletes from their race walking. Projection plane defined by the first two *2-PCs* of second stage *2-PCA*. Different colored clusters represent different race walkers.

In agreement with results presented by Das et al. [29, 30], the two-stage principal component analysis revealed to be a useful instrument for representing different motor behaviors. In fact the "O"-shaped manifold provided a canonical low-dimensional representation of a large amount of race walking variables. It captured the majority of data variance and provided an optimal way for describing spatio-temporal relationships among gait variables. Unfortunately, *2-PCA* results were hardly interpretable in an intuitive way.

Inter-joint coordination patterns were further analyzed through bivariate *f-PCA*. This statistical technique had the advantage to present results resembling those derived from standard coordinative analyses. Moreover, it evaluated simultaneously only two variables, thus improving the readability of results. Two variables were simultaneously evaluated with *f-PCA*. The attention was focused on the relations between two consecutive joints: hip and knee, knee and ankle. Bivariate functional principal components were named *bf-PCs*.

Hip-knee coordination

The individual components for the hip and knee and the bivariate components were plotted respectively in parts (d),(a) and (b) of Fig.5.28. Moreover *bf-PCs* scores were inserted in a scatterplot shown in part (c). Similar representations will be adopted for *fb-PC1* and *fb-PC2* in all the following coordination analyses.

The first five *bf-PCs* calculated from the hip and knee data accounted for more than 90% of the total variation of hip-knee coordination (Fig.5.27). In particular *bf-PC1* explained most (about 63.1%) of the total variation.

Mean hip and knee joint angles are colored in different ways according to the race walking phase: red for SP, green for FSP, cyan for RSP. Swing phase lasted almost half of the entire race walking cycle; moreover from FSP to RSP the movement was described by an increased number of frames, thus revealing a faster movement while passing through the vertical position.

Inspection of Fig.5.28(b) revealed that *bf-PC1* corresponded to the variation around the mean hip-knee curve. Arrows pointed approximately 45° relative to the horizontal axis with an upward direction. This suggested that athletes who scored positively on *bf-PC1* were characterized by an action that involved hip and knee joint flexion more than the mean curve throughout the race walking cycle. This behavior was adopted by {s1, s5, s7} (Fig.5.28 (c)), while {s2, s3, s4, s6}, with negative scores, presented greater hip and knee extension throughout the movement. Once again s1 and s6 showed the most clearly separated clusters of trials, in agreement with Section 5.2. Thus the

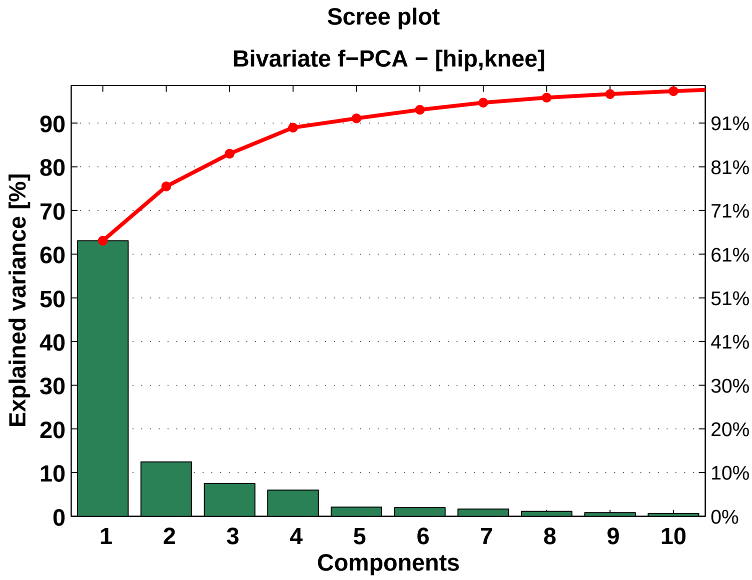


Figure 5.27: Scree plot showing the percentage of variation in the data explained by the functional principal components in the bivariate analysis of hip and knee angles.

interaction between hip and knee movement revealed to be an essential factor for athletes' skill discrimination.

Inspection of Fig.5.29(b) reveals that *bf-PC2* caused less homogeneous variations for hip and knee joint angles respect to the mean curve during the three phases. Race walkers with high positive scores of *bf-PC2* used to increase the knee and the hip extension during the early swing. This movement was followed by an higher knee flexion in the FSP, a greater extension in the early RSP and then again a flexion in the last RSP.

Thus athletes {s2, s4, s7}, with positive scores on *bf-PC2* (Fig.5.29(c)) tended to have an increased knee flexion during the passage through the vertical position, so they may incur in the violation of the rule of locked knee. Moreover, s6 had high negative scores both for *bf-PC1* and *bf-PC2* and was characterized by a great hip and knee extension (hyperextension) at heel-strike. Since this athlete was included among those who may incurred in lower knee ligaments injuries in the previous univariate analysis (Section 5.2), this result could suggest that particular attention should be paid to prevent any knee damage.

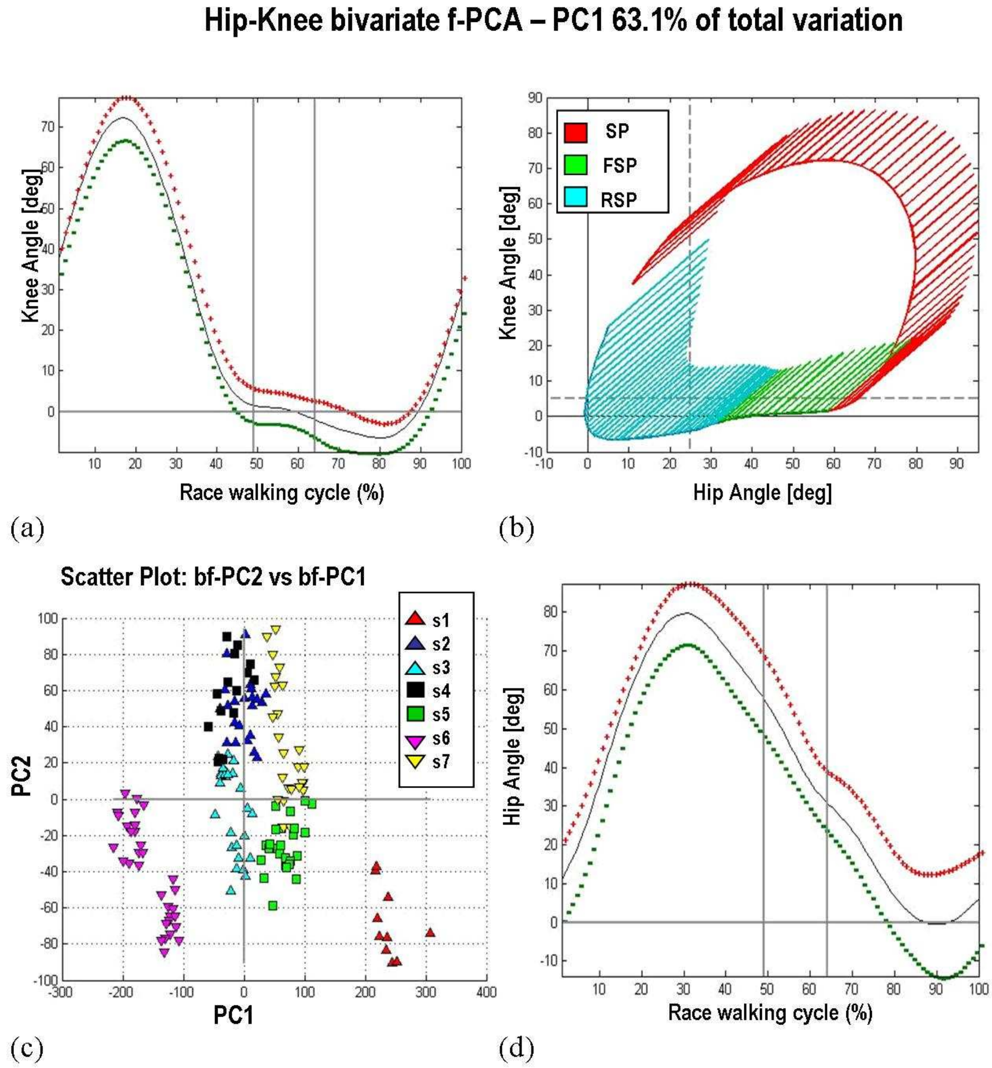


Figure 5.28: Bivariate functional *PC1* for hip-knee joint coordination: (a,d) deviation from the mean curve due to hip and knee *bf-PC1* respectively; (b) mean hip-knee angle-angle plot with the contribution of *bf-PC1* represented as vectors; (c) *bf-PC2* vs *bf-PC1* scores scatterplot. Horizontal and vertical dotted lines in part (a) are approximatively the mean angular values during standing.

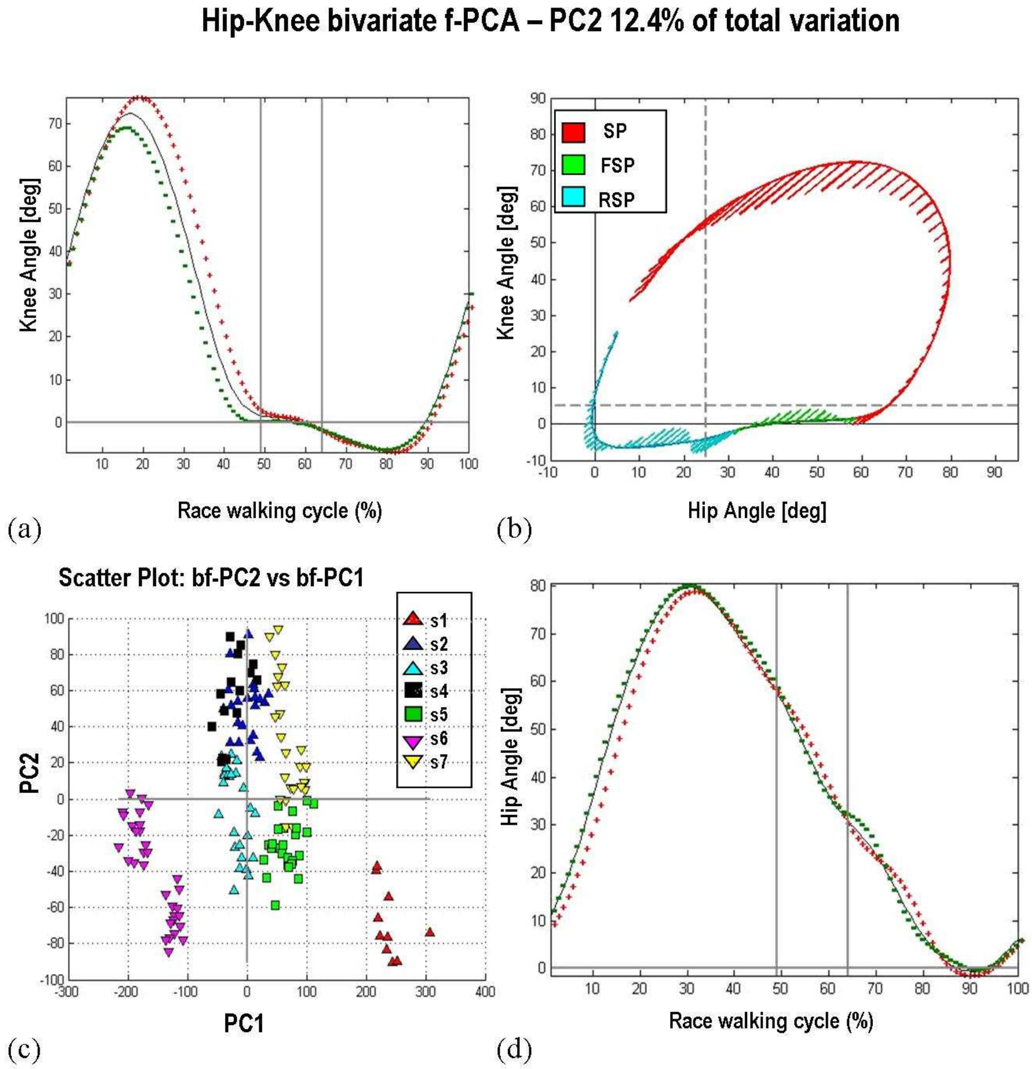


Figure 5.29: Bivariate functional *PC2* for hip-knee joint coordination: (a,d) deviation from the mean curve due to hip and knee *bf-PC2* respectively; (b) mean hip-knee angle-angle plot with the contribution of *bf-PC1* represented as vectors; (c) *bf-PC2* vs *bf-PC1* scores scatterplot. Horizontal and vertical dotted lines in part (a) are approximatively the mean angular values during standing.

Knee-ankle coordination

Analysis of coordination was then carried out on the knee and angle angles. The first five *bf-PCs* accounted for more than 92% of data variance (Fig.5.30). The first bivariate functional component was the most representative one.

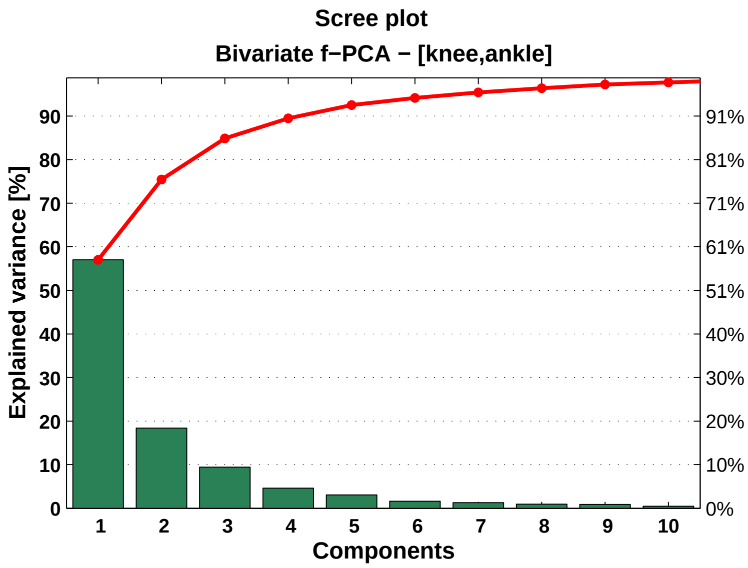


Figure 5.30: Scree plot showing the percentage of variation in the data explained by the functional principal components in the bivariate analysis of knee and ankle angles.

Inspection of Fig.5.31(b) revealed that *bf-PC1* corresponded to an overall coordinative deviation from the mean trend throughout the race walking cycle. Race walkers with high positive scores on *bf-PC1* for knee-ankle coordinative joints, used to increase the knee joint flexion and the ankle plantarflexion. This behavior was particularly remarkable during the swing phase. In agreement with previous results, s6, scoring negatively on *bf-PC1* (Fig.5.31(c)), showed a tendency to hyperextend the knee (Fig.5.31(b)). This movement implied a more plantarflexed ankle (Fig.5.31(d)).

Fig.5.32(b) represented the influence of *bf-PC2* on the mean coordinative curve. For race walkers with positive *bf-PC2* scores (Fig.5.32(c)) a large increase in ankle dorsiflexion and knee extension is well recognized during the late swing. This indicated that {s1, s3, s5} correctly performed the request knee extension.

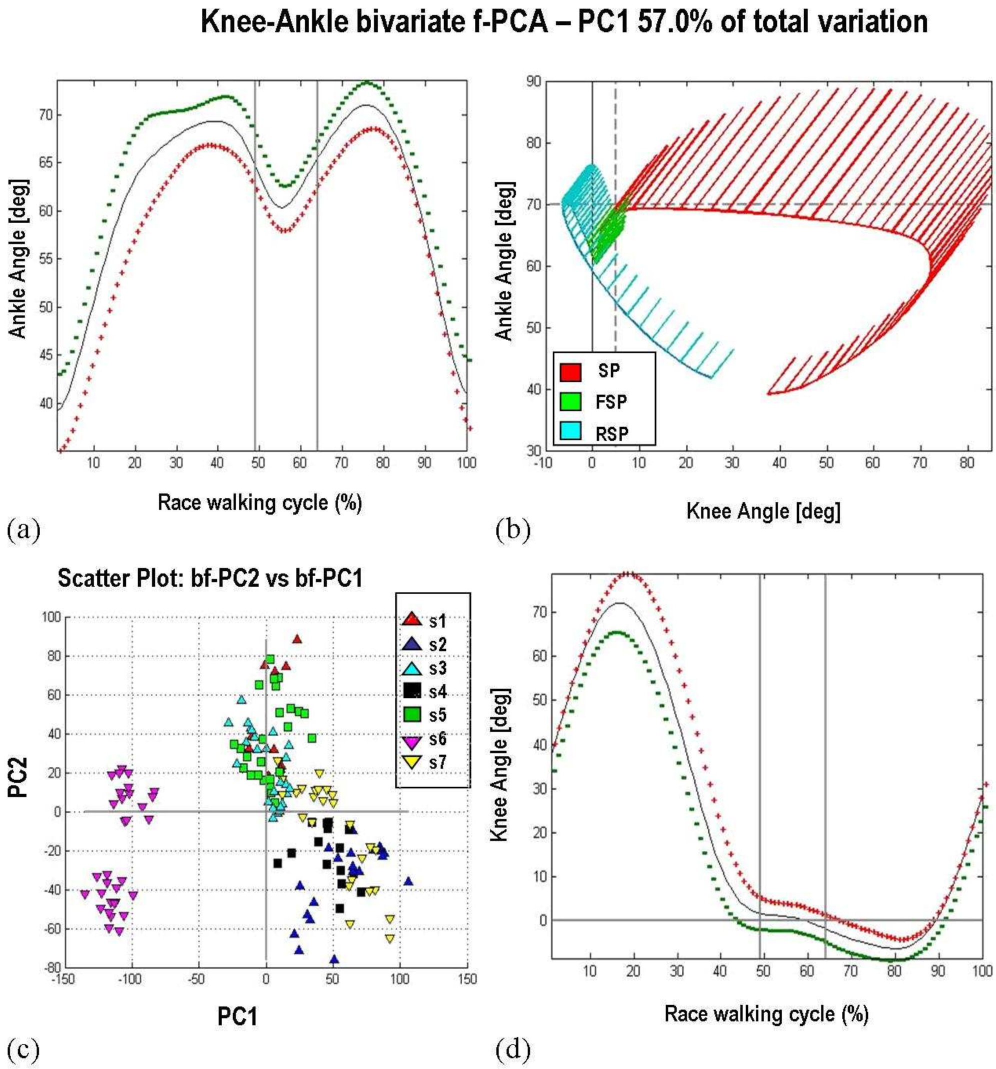
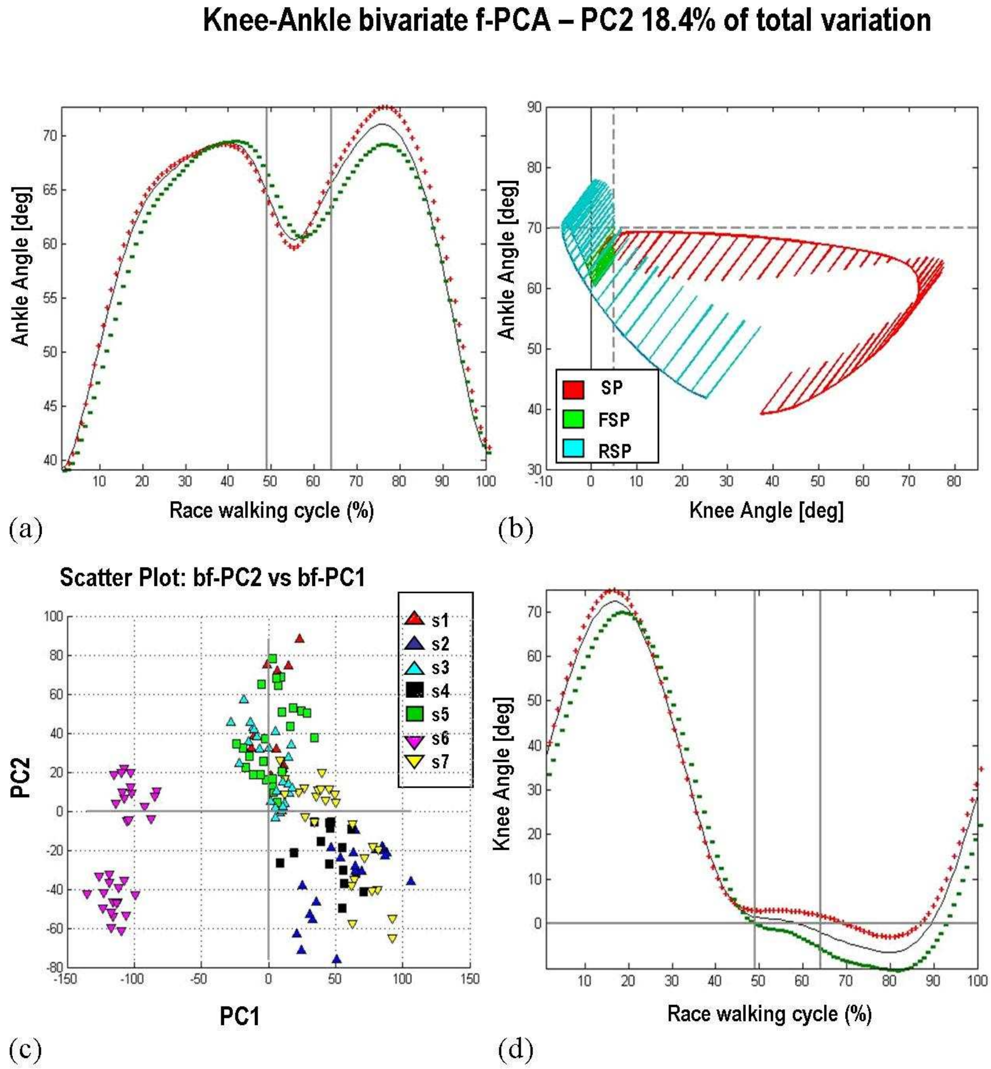


Figure 5.31: Bivariate functional PCA of *bf-PC1* for knee-ankle joint coordination: (a,d) deviation from the mean curve due to knee and ankle *bf-PC1* respectively; (b) mean knee-ankle angle-angle plot with the contribution of *bf-PC1* represented as vectors; (c) *bf-PC2* vs *bf-PC1* scores scatterplot. Horizontal and vertical dotted lines in part (a) are approximatively the mean angular values during standing.



(a) (b) (c) (d)

Figure 5.32: Bivariate functional PCA of *bf-PC2* for knee-ankle joint coordination: (a,d) deviation from the mean curve due to knee and ankle *bf-PC2* respectively; (b) mean knee-ankle angle-angle plot with the contribution of *bf-PC2* represented as vectors; (c) *bf-PC2* vs *bf-PC1* scores scatterplot. Horizontal and vertical dotted lines in part (a) are approximatively the mean angular values during standing.

5.4 Longitudinal monitoring

Data acquired in two different sessions (session *I* and session *II*) were compared. This analysis allowed to evaluate PCA potential in following athletes during a longitudinal monitoring. Four subjects repeated the analyses: s1, s2, s3 and s6. Results of s6 (group3) were taken into consideration as an explanatory example. In fact, this athlete presented the most evident differences between the two sessions.

f-PCA was applied to each variable for data resulting from sessions *I* and *II*. Scores of the two sessions were then compared using boxplots. This technique was similar to that adopted by Ryan and Harrison [51, 100] to compare different stages in the development of the vertical jump for young subjects. While in vertical jump analysis boxplots were applied to indicate changes on scores of children at different developmental stages, in this thesis they were used on scores of the same subject in different training periods. Hence, they helped to longitudinal monitor athlete's changed behaviors.

A biomechanical interpretation of the differences in scores was inferred from the analysis of figures and results reported in Section 5.2. The representation of the influence of *f-PCs* on the mean trend were inserted another time in the figures, to improve their immediate inspection.

For some kinematic variables (pelvis rotation, hip and knee joint angles in the sagittal plane), *f-PC1* scores became less variable and more positive in the passage from session *I* to session *II*, to the detriment of ankle joint ankle, which presented a larger range in scores values (Fig.5.33 - Fig5.36).

Reduced variability in *f-PC1* and *f-PC2* scores for pelvis rotation (Fig.5.33 (a)) showed a developmental trend. In fact, pelvis internal-external rotations became less variable, probably because the athlete was trained to maintain a correct movement in attenuating the vertical *COM* oscillations.

The more positive score values on *f-PC1* for hip and knee demonstrated a tendency of this athlete to maintain these joints more flexed throughout the race walking cycle (Fig.5.34 (b)). This behavior might be interpreted as an improvement due to training. In fact, in previous analyses (Section 5.2) hip flexion-extension revealed to be a fundamental aspect in revealing athlete's ability. Moreover, s6 showed a knee hyperextension that might put stress on the posterior structures of the knee joint.

The less variability of functional component scores, evidenced a reduced knee extension in session *II* (Fig.5.35 (b)). This behavior might be a movement trained to prevent injuries on this joint.

The larger range of scores on *f-PC1* for the ankle angle in session *II* might be imputed to a compensation movement in learning a new motor skill. In fact, a more variable behavior at the ankle dorsi- plantarflexion could facilitate changes in coordination of hip and knee. It might be interpreted as a functional variability induced by more restricted constraints for hip and knee joints.

More positive scores were evident for hip and knee angles also on the second functional principal component. Section 5.2 revealed that *f-PC2* concerned a better hip and knee behavior during the passage through the vertical position (Fig.5.34 (c), Fig.5.35 (c)). Hence, boxplots suggested that the increase in the first and second functional principal component for hip and knee angles could be developmentally related.

The lower variability in score values, especially for the hip and knee joints, might be also imputed to a more symmetrical behavior, limiting the differences in right and left limbs demonstrated in Section 5.2.

The improvement of knee angle flexion-extension behavior was clearly visible in Fig.5.41. Clusters of trials belonging to right and left limbs were no more divided passing from session *I* to session *II*.

The scores of *f-PC2* for the ankle angle on the sagittal plane became more positive. This corresponded to an increased ankle plantarflexion after the passage through the vertical position (Fig.5.36 (c)). This behavior might help the athlete in the toe-off phase and might facilitate to maintain the knee in the extended position for a longer time, as previously described in Section 5.2.

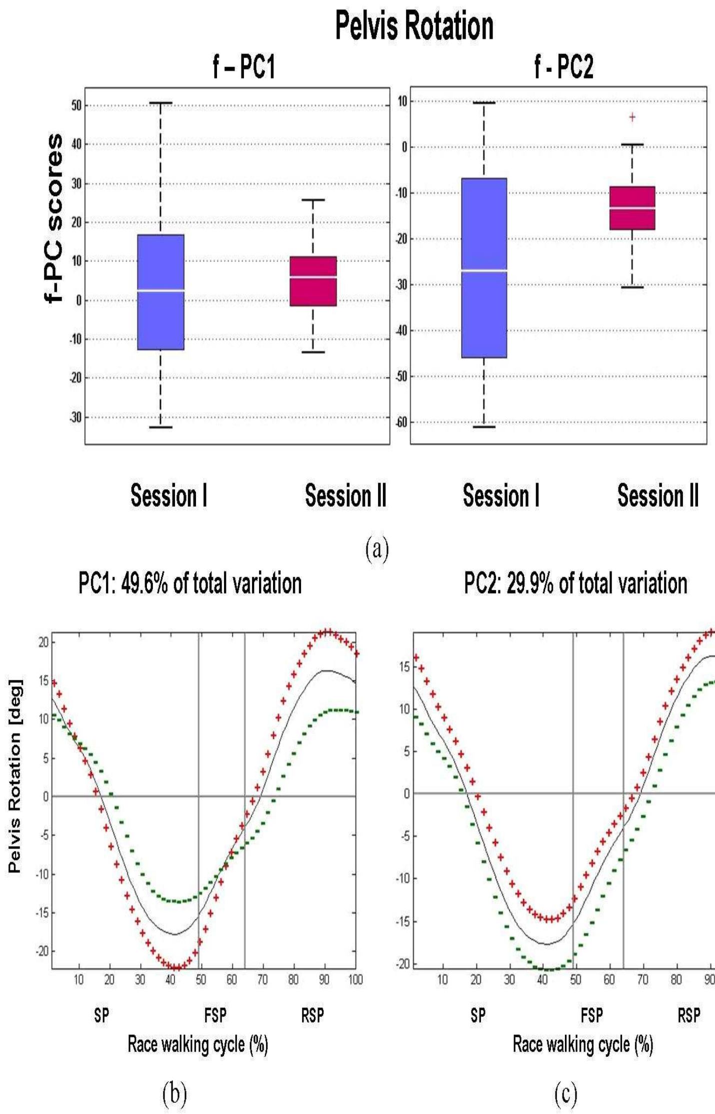


Figure 5.33: (a) Boxplot: scores of the *f-PC1* and *f-PC2* for pelvis rotation in the sagittal plane, divided in two groups, corresponding to sessions I and II. (b,c) The mean pelvis rotation curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c).

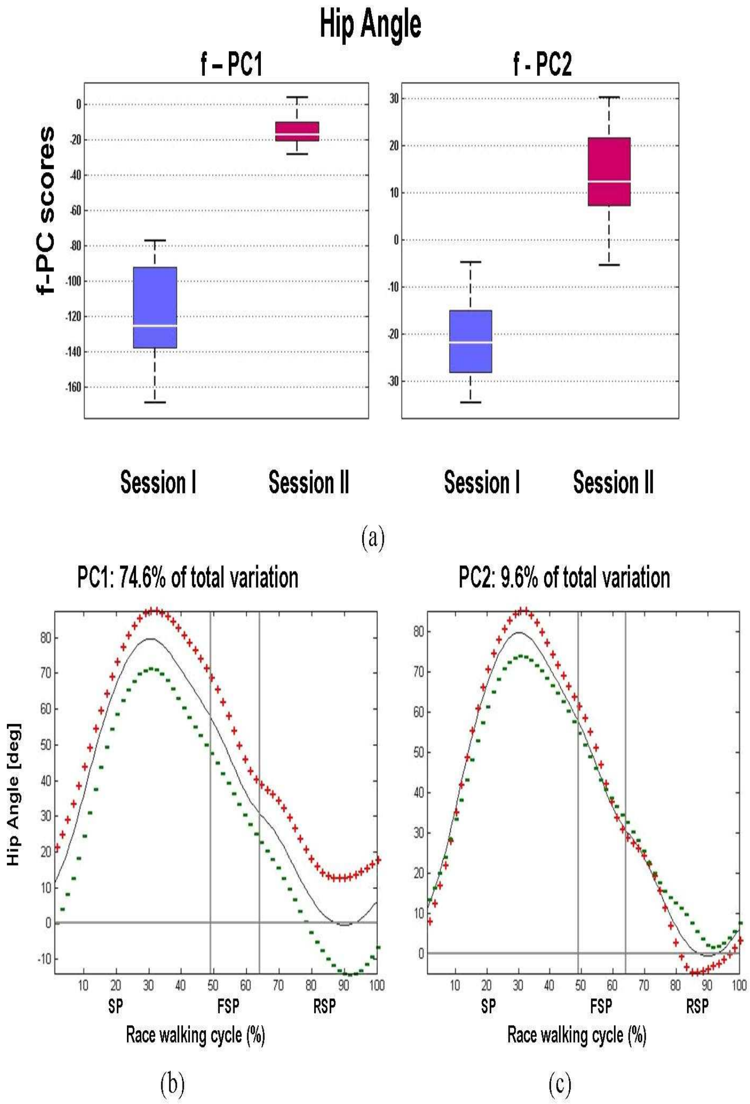


Figure 5.34: (a) Boxplot: scores of the *f-PC1* and *f-PC2* for hip joint angle in the sagittal plane, divided in two groups, corresponding to sessions I and II. (b,c) The mean hip joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c).

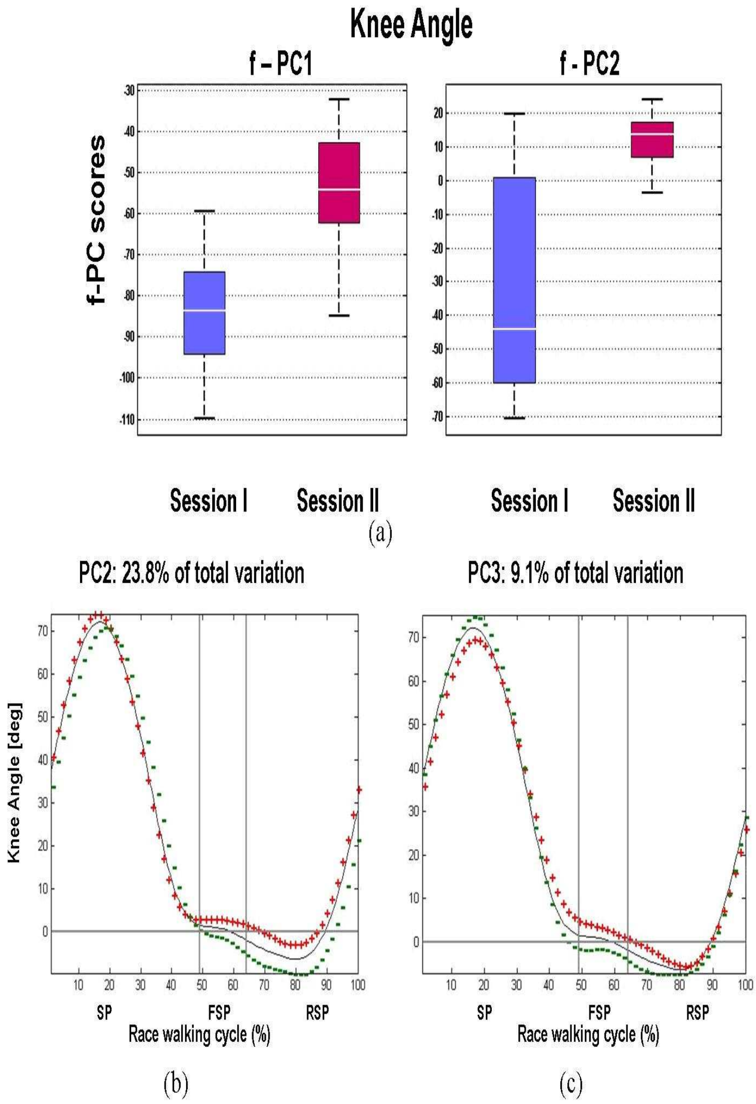


Figure 5.35: (a) Boxplot: scores of the *f-PC1* and *f-PC2* for knee angle in the sagittal plane, divided in two groups, corresponding to sessions *I* and *II*. (b,c) the mean knee joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c).

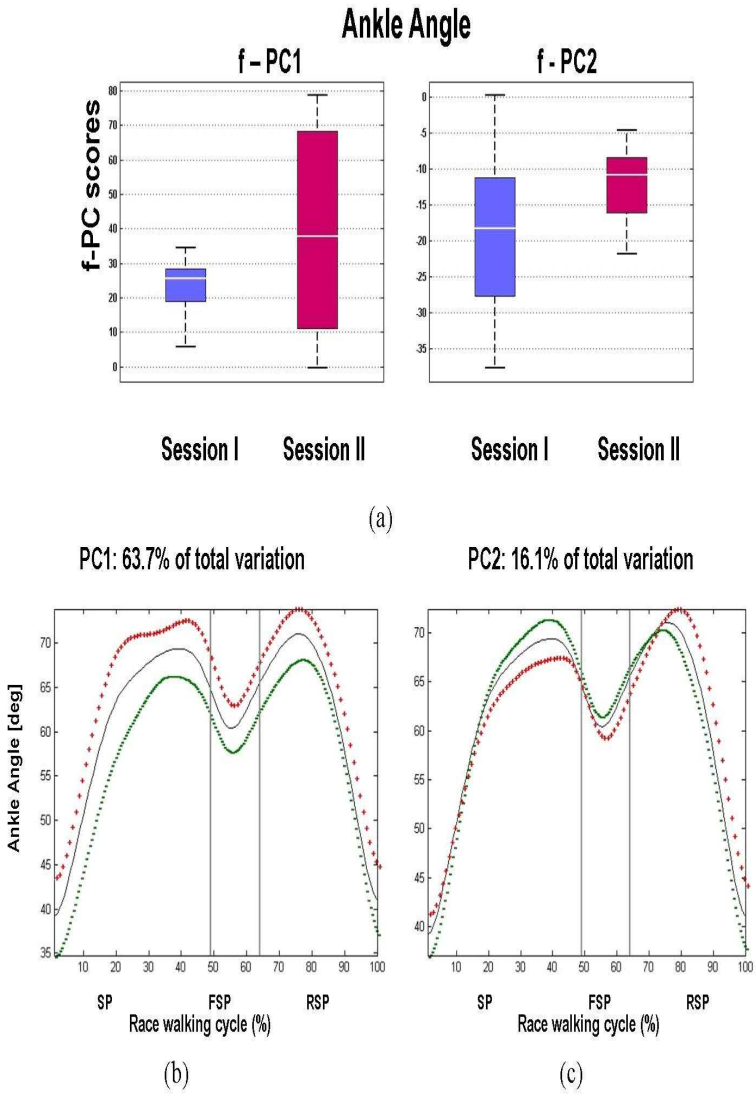


Figure 5.36: (a) Boxplot: scores of the *f-PC1* and *f-PC2* for ankle joint angle in the sagittal plane, divided in two groups, corresponding to sessions *I* and *II*. (b,c) the mean ankle joint angle curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c).

The kinetic variables (Fig.5.37 - Fig.5.40) showed a similar trend, even if the differences in scores' spread were less evident. Mediolateral and vertical GRF showed a higher reliability in performing the movement. In fact, ranges of score values reduced from session *I* to session *II*.

Moreover, more positive scores on *f-PC2* for mediolateral GRF showed an increased delay (Fig.5.37 (c)) in the movement. This characteristic belonged to the best performing athletes, as described in Section 5.2. Hence, this was a new proof of the athlete's improvement in maintaining a straight progress of *COM*.

The vertical GRF showed no evident differences between session *I* and session *II* (Fig.5.38 (a)). Less variable scores described a more reproducible movement among repeated trials.

The knee joint moment presented a decrease in scores variability, especially for *f-PC2* (Fig.5.39 (a)). This result might be related to the correction of the asymmetric moments of right and left knees, outlined in Section 5.2. Moreover, the more positive score values of knee moment for *f-PC1* and *f-PC2* were strictly related to the improvement of knee flexion extension movement in session *II*.

In agreement with previous results, ankle joint moments showed a larger range of score values (Fig.5.40 (a)). More negative scores for both *f-PC1* and *f-PC2* revealed a delayed and stronger plantarflexion moment just before passing through the vertical position (Fig.5.40 (b), Fig.5.40 (c)).

A further immediate evaluation of athletes' performance improvement caused by training derived by the application of *2-PCA* to data obtained for all the four athletes who repeated the acquisition sessions.

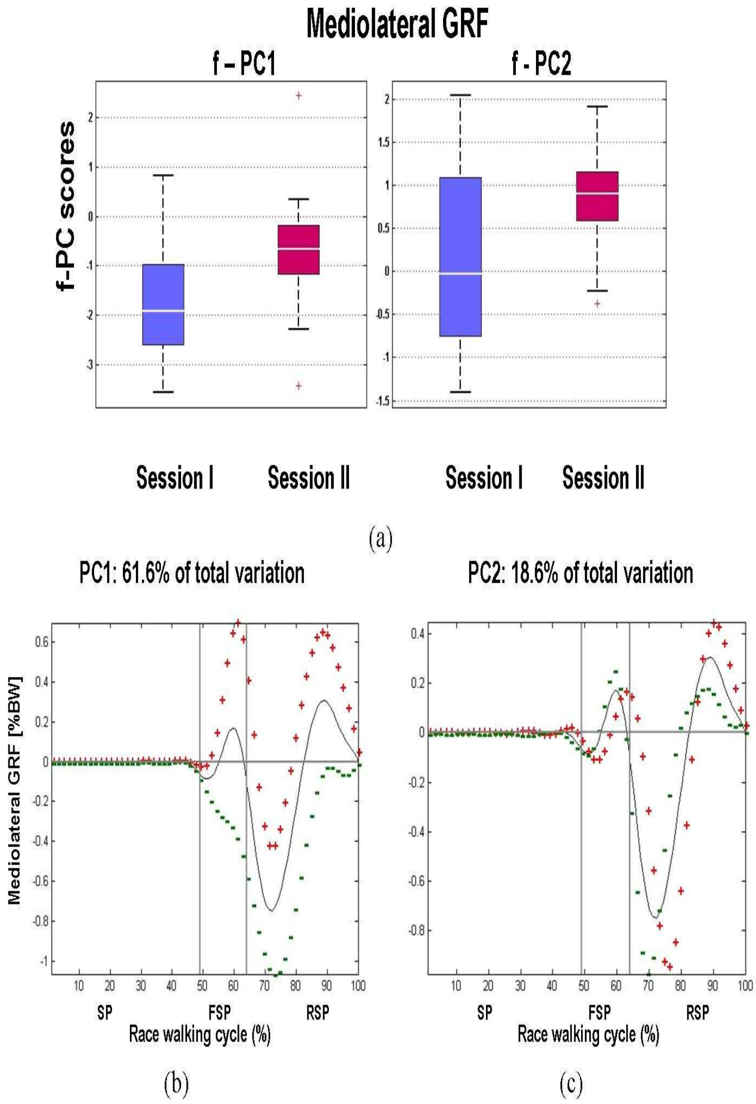


Figure 5.37: (a) Boxplot: scores of the *f-PC1* and *f-PC2* for mediolateral GRF, divided in two groups, corresponding to sessions *I* and *II*. (b,c) the mean mediolateral GRF curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c).

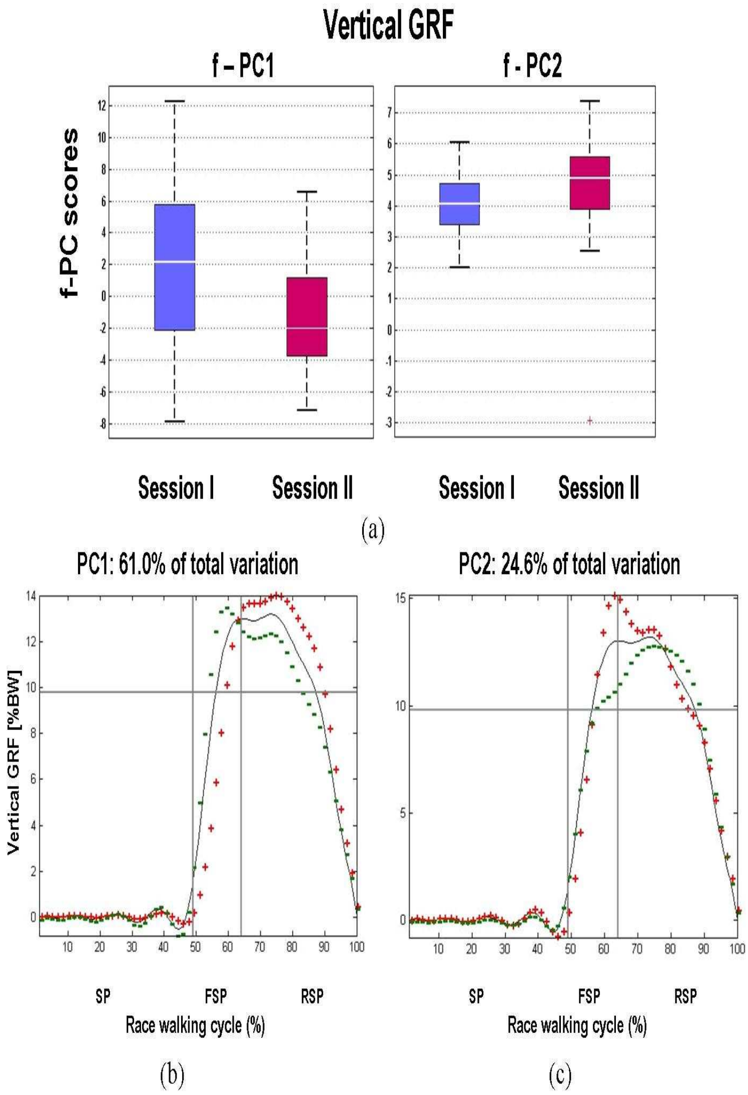


Figure 5.38: (a) Boxplot: scores of the *f-PC1* and *f-PC2* for mediolateral GRF, divided in two groups, corresponding to sessions *I* and *II*. (b,c) the mean vertical GRF curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c).

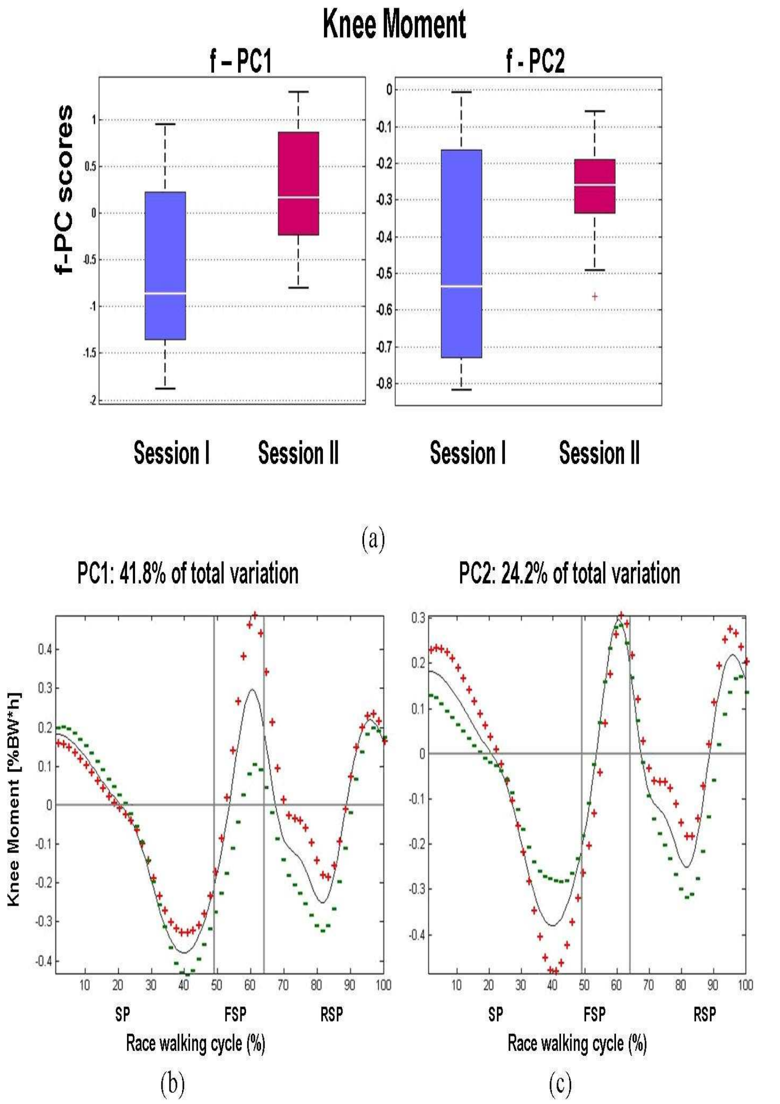


Figure 5.39: (a) Boxplot: scores of the *f-PC1* and *f-PC2* for knee joint moment in the sagittal plane, divided in two groups, corresponding to sessions *I* and *II*. (b,c) the mean knee joint moment curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c).

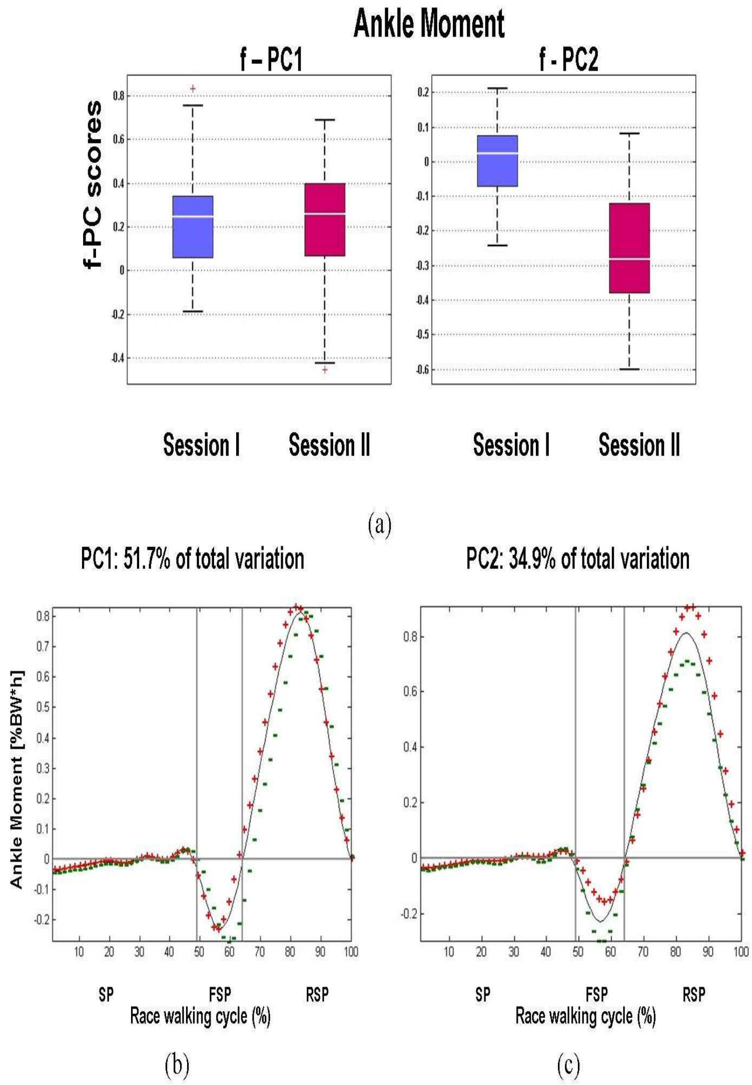


Figure 5.40: (a) Boxplot: scores of the *f-PC1* and *f-PC2* for ankle joint moment in the sagittal plane, divided in two groups, corresponding to sessions *I* and *II*. (b,c) the mean ankle joint moment curve is shown with curves created by adding (red +) and subtracting (green -) a multiple of: *f-PC1* (b), *f-PC2* (c).

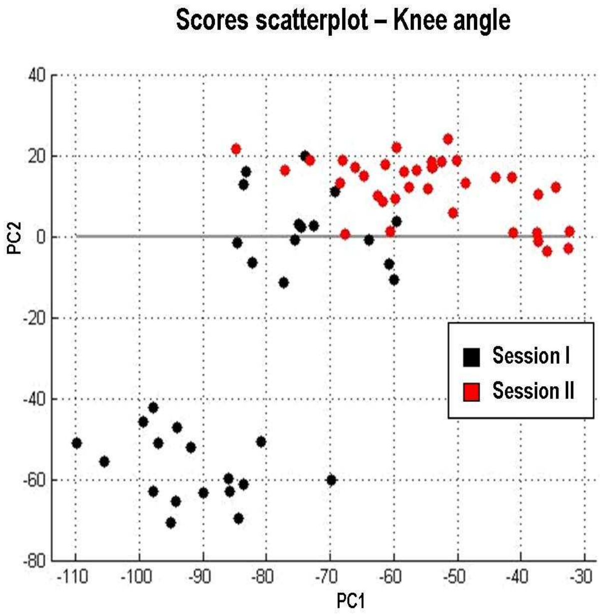


Figure 5.41: Scores scatterplot of the *f-PC1* and *f-PC2* for knee joint angle in the sagittal plane for s6; data are divided in two groups, corresponding to sessions *I* and *II*.

Results from the first stage *2-PCA* for s6 were reported in Fig.5.42. Interestingly, Fig.5.42 resembled Fig.5.25, where two athletes, belonging to different skill levels, were compared. The manifold of s6 in the session *II* was very similar to the one of s3 (group1) for session *I*. It seemed that the improvement of s6 from session *I* to session *II* changed the race walking movement, leading it to be more similar to the one of the most performing athletes. In fact, in session *II*, s6 had a more smooth movement from the late swing to the first RSP (positive projections on *2-PC1*. Moreover, data points were less scattered among different trials, especially during the swing phase. In early SP the differences in trials acquired for different limbs (right and left side), evident in session *I*, were no more visible in the second one. It implied an higher symmetry in lower limbs movement. This results confirmed an improvement in motion repeatability, already seen with boxplots analysis.

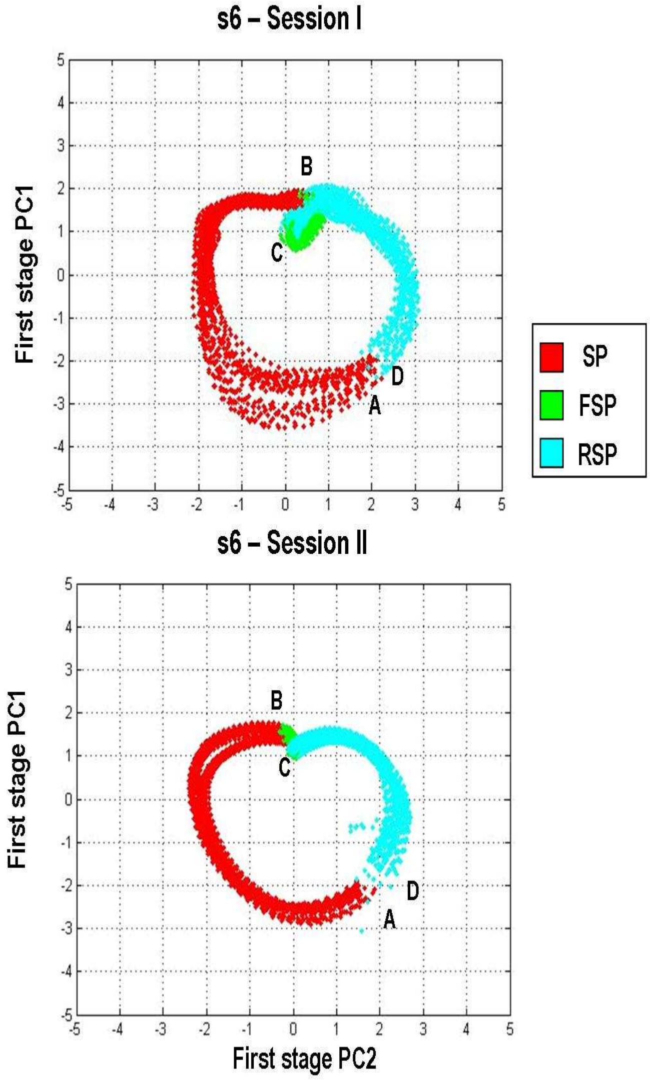


Figure 5.42: Longitudinal monitoring: manifolds in (*2-PC1*, *2-PC2*) space after first stage of PCA for s6 resulting from the acquisition session I (a) and II (b).

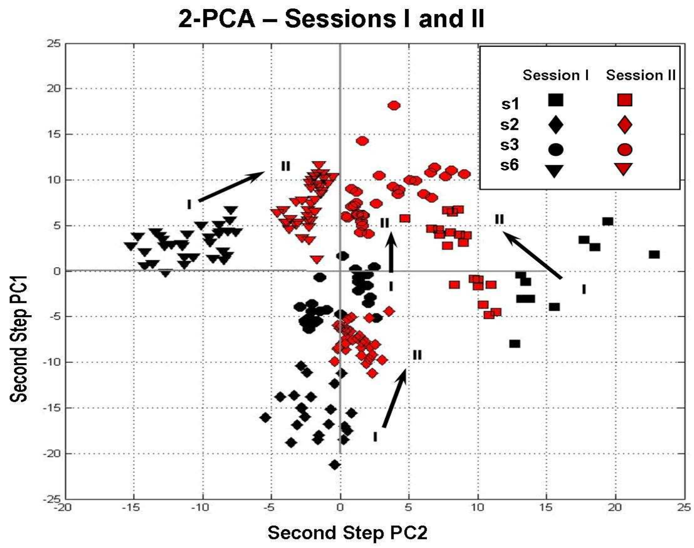


Figure 5.43: Scores scatterplot for second stage *2-PCA* (*2-PC1* vs *2-PC2*); comparison between sessions I and II for s1, s2, s3 and s6

The analysis of the scores of the second stage *2-PCA* (Fig.5.4) indicated a tendency for *2-PC1* and *2-PC2* scores of all the four race walkers toward the first quadrant. This might be an indication of a general improvement caused by training in all the athletes. These results would be probably hidden by biovariability in a standard gait analysis evaluation.

To resume, PCA method provided specific information on the population sampled. This technique revealed useful in understanding the differences among athletes of different skill levels as well as for carrying out a longitudinal monitoring.

Using principal component analysis, race walking waveform data were represented as a set of scores and components that provided important information. Scores scatterplot gave immediate visual evidence of main differences among athletes. As such, they were a suitable choice due to their simplicity and ease of interpretation. Moreover, principal components allowed the interpretation of these differences by identifying the portion of the race walking cycle in which they occurred.

The investigation of single variables evidenced how differences detected by the principal component techniques were correlated not only with skill level, but also with the peculiar characteristics of each athlete. For example, the best athletes (group1), with similar competitive results, were associated by similar hip flexion extension movements. Meanwhile, they demonstrated quite different behaviors at the knee joint angle.

Important coordinative information were inferred both by *f-PCA* and *2-PCA*. While in the first case results were easily understandable, resembling those derived from standard coordinative methods, the second statistical technique was hardly comprehensible. In fact, applying twice PCA on data, created a gap between final results and the original variables.

Concerning the longitudinal monitoring, functional principal component scores indicated a developmental trend in the first and second components, especially for the kinematic variables. They represented a reduction in variability, indicating that asymmetries were limited and a more repeatable pattern was adopted.

Conclusions

Much research in sports biomechanics have been based on particular assumptions. One of them is the presumption of the existence of an optimal movement model that can be applied to all athletes. An other assumption is that one trial of a movement task may represent also the other trials. These hypotheses are hardly true in most of the experimental cases. In fact, data are always affected by inter- and intra-subject variability. Moreover, most of the approaches in sports research are based only on performance results, rather than studying the whole movement process. Hence, there is the need to explore innovative approaches for measuring and assessing movement analysis data.

This study has shown that PCA can be applied to sports studies of kinematic and kinetic data. The key concepts, techniques and advantages of using PCA were illustrated using an analysis of race walking. Three different approaches were used, to inspect all the potential applications of PCA: traditional (*t-PCA*), functional (*f-PCA*) and to-stage (*2-PCA*).

First, a comparison between *t-PCA* and *f-PCA* methods was carried out. The first technique consisted in the application of PCA to data described on a discrete temporal grid, while the second one corresponded more to the functional nature of data. Even if *t-PCA* needed a greater number of principal components to explain the same amount of variance, results did not present evident differences between the two techniques. This was probably due to the intrinsic nature of data. Since anomalous and temporally odd curves have already been removed in the preprocessing phase, the smoothing adopted by *f-PCA* had no major effect. Moreover kinematic and kinetic data were collected with quite high frequency (respectively 100Hz and 500Hz). Thus, the discrete temporal grid was dense and discrete data were comparable with their functional representations. Hence, we might infer that traditional PCA is preferred to be adopted on data with regular patterns and collected with high frequency. Otherwise, functional PCA, even if implying some efforts in smoothing data, should be used.

A biomechanical description of race walking and a general characterization of all the athletes were carried out. The applied methods showed specific advantages in solving different challenges. *f-PCA* allowed to identify the most powerful variables able to discriminate athletes belonging to different skill levels. Scatterplots of functional component scores clustered trials of different subjects in clearly divided groups for most of the analyzed variables. Moreover, *2-PCA* identified the main "motion features" that characterize race walking. Hence, the most representative variables for the description of performance, together with their mutual relations, were found. Therefore, these analytical techniques proved high accuracy for identity discrimination tasks. They yielded also insights into the contributions of various parameters to race walking principal modes.

A coordinative description of race walking performance was also inferred for every athlete through *bf-PCA* and *2-PCA*. The first technique allowed to obtain results resembling those derived by continuous phasing relationships (CRP). Hence, a straightforward biomechanical interpretation was possible. While *bf-PCA* allowed analyses conceiving only two joints, *2-PCA* simultaneously evaluated a larger number of variables. It revealed to be a very powerful statistical technique in underlying differences in coordinative behaviors among athletes.

Despite this promising recognition potentiality in classifying race walkers, *2-PCA* presented hardly interpretable results. In fact, difficulties were encountered in understanding the meaning of the manifolds and in associating the second stage results to the original variables. Perhaps, the set of variables inserted in the analysis influences the comprehension of results. Probably, including more variables or changing their choice, might improve the performance of this technique.

An example of longitudinal monitoring of an athlete was also reported in the thesis. Functional scores for kinematic variables proved to change drastically from session *I* to session *II*. One of the variables that mostly modified its behavior in the two periods of training season was hip joint angle in the sagittal plane. Throughout the above-mentioned analyses, hip flexion extension movement demonstrated to be a fundamental aspect for race walking skill characterization. Hence, results of the longitudinal monitoring were interpreted as a consequence of training. New evidence of this improvement derived from the evaluation of the two *2-PCA* manifolds of the athlete in the two sessions. Asymmetries between right and left limbs were clearly reduced and a smoother movement was performed.

To resume, PCA provides a basis for sports biomechanics, for ensuring that important data are not sacrificed in the analysis and that important trends in the kinematic and kinetic patterns are not missed through limitations in the statistical analysis procedure. PCA revealed to represent an important mean of performance analysis and skill characterization. Athletes' peculiar motor strategies were discovered even for subjects associated by similar competition results. Probably such subtle characteristics might be hardly found through traditional data analysis techniques. An example of how PCA might be used in longitudinal monitoring of athletes, showed how this technique might quantitatively support individual training procedures.

Perhaps, this statistical technique might be adopted also in seeking to reduce the incidence of injuries, giving indication of individual possible dangerous behaviors. Further investigations are needed to support this hypothesis. Despite many interesting PCA potentialities, further efforts should be spent in improving the interpretation of results and in finding a way to make them intelligible for practitioners.

In a longtime perspective, PCA could be applied on other kind of movement tasks and the proof of its advantages might be strengthened. Moreover, inferred information could be inserted in a graphical interface to present an immediate feed-back of the athletes peculiarities, thus helping to achieve the required outcome of improved performance.

Bibliography

[1] H. Ahn, E. Choi, and I. Han. Extracting underlying meaningful features and canceling noise using independent component analysis for direct marketing. *Expert Systems with Applications*, 33:181–191, 2007.

[2] J.L. Astephen and K.J. Deluzio. Changes in frontal plane dynamics and the loading response phase of the gait cycle are characteristic of severe knee osteoarthritis application of a multidimensional analysis technique. *Clinical Biomechanics*, 20:209–217, 2005.

[3] S. Barker, R. Craik, W. Freedman, N. Herrmann, and H. Hillstrom. Accuracy, reliability, and validity of a spatiotemporal gait analysis system. *Med Eng Phys*, 28(5):460–467, 2006.

[4] V. Barra. Analysis of gene expression data using functional principal components. *Computer Methods and Programs in Biomedicine*, 75, 2004.

[5] R. Bartlett. Future trends in sports biomechanics - reducing injury risk or improving performance? In *Proceedings of the XXIII ISBS Symposium*, 2005.

[6] R. Bartlett. Artificial intelligence in sports biomechanics: new dawn or false hope? *Journal of Sports Science and Medicine*, 5(443):474–479, 2006.

[7] R.M. Bartlett. Current issues in the mechanics of athletic activities. a position paper. *Journal of Biomechanics*, 30:477–486, 1997.

[8] R.M. Bartlett, J. Wheat, and M. Robins. Is movement variability important for sports biomechanics? *Sports Biomechanics*, 4(2):224–243, 2007.

[9] J.G. Barton and A. Lees. An application of neural networks for distinguishing gait patterns on the basis of hipknee joint angle diagrams. *Gait and Posture*, 5:28–33, 1997.

[10] R. Begg and J. Kamruzzaman. A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *Journal of Biomechanics*, 38:401–408, 2005.

[11] E. Bingham. *Advances in independent component analysis with applications to data mining.* PhD thesis, Helsinki University of Technology, 2003.

[12] W. Blajer and W. Schielen. Walking without impacts as a motion/force control problem. *ASME Journal of Dynamic Systems, Measurement and Control*, 114, 1992.

[13] M.A. Cairns, R.G. Burdett, J.C. Pisciotta, and S.R. Simon. A biomechanical analysis of racewalking gait. *Medicine and Science in Sports and Exercise*, 18(4):446–453, 1986.

[14] A. Cappello and R. Stagni. *Errori e artefatti in stereofotogrammetria e loro compensazione*, pages 103–129. Patron, 2003.

[15] R.B. Cattell. The scree test for the number of factors. *Multivariate Behavioral Research*, 1:245–276, 1966.

[16] M.M. Cerney. *From gesture recognition to functional motion analysis: Quantitative techniques for the application and evaluation of human motion.* PhD thesis, Iowa State University, 2005.

[17] T.H. Chang and Y. Hurmuzlu. Sliding control without reaching phase and its application to bipedal locomotion. *ASME Journal of Dynamic Systems, Measurement and Control*, 105, 1994.

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

[19] T. Chau and S. Redekop. Managing variability in the summary and comparison of gait data. *J Neuroengineering Rehabil*, 2(22), 2005.

[20] T. Chau, S. Young, and S. Redekop. Managing variability in the summary and comparison of gait data. *Journal of NeuroEngineering and Rehabilitation*, 2(22), 2005.

[21] V.L. Chester and A.T. Wrigley. The identification of age-related differences in kinetic gait parameters using principal component analysis. *Clinical Biomechanics*, doi:10.1016/j.clinbiomech.2007.09.007, 2007.

[22] L. Chiari, U. Della Croce, A. Leardini, and A. Cappozzo. Human movement analysis using stereophotogrammetry. part 2: instrumental errors. *Gait and Posture*, 21(2):197–211, 2005.

[23] G. Courtine and M. Schieppati. Tuning of a basic coordination pattern constructs straight-ahead and curved walking in humans. *Journal of Neurophysiology*, 91:1524–1535, 2004.

[24] U. Della Croce, A. Cappozzo, and D.C. Kerrigan. Pelvis and lower limb anatomical landmark calibration precision and its propagation to bone geometry and joint angles. *Med Biol Eng Comput*, 37(2):155–161, 1999.

[25] U. Della Croce, A. Leardini, L. Chiari, and A. Cappozzo. Human movement analysis using stereophotogrammetry. part 4: assessment of anatomical landmark misplacement and its effects on joint kinematics. *Gait and Posture*, 21(2):226–237, 2005.

[26] A. Daffertshofer, C.J.C. Lamoth, O.G. Meijer, and P.J. Beek. Pca in studying coordination and variability: a tutorial. *Clinical Biomechanics*, 19:415–428, 2004.

[27] M. D'Amico and G. Ferrigno. Technique for the evaluation of derivatives from noisy biomechanical displacement data using a model-based bandwidth-selection procedure. *Med Biol Eng Comput*, 28:407–415, 1990.

[28] S. Das, M. Lazarewicz, and L.H. Finkel. Principal component analysis of temporal and spatial information for human gait recognition. In *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, 2004.

[29] S.R. Das, R.C. Wilson, M.T. Lazarewicz, and L.H. Finkel. Gait recognition by two-stage principal component analysis. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, 2006.

[30] S.R. Das, R.C. Wilson, M.T. Lazarewicz, and L.H. Finkel. Two-stage pca extracts spatiotemporal features for gait recognition. *Journal of Sports Sciences*, 1(5):9–17, 2006.

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

[32] K.J. Deluzio, U.P. Wyss, P.A. Costigan, C. Sorbie, and B. Zee. Gait assessment in unicompartmental knee arthroplasty patients: Principal component modelling of gait waveforms and clinical status. *Human Movement Science*, 18(5):701–711, 1997.

[33] K.J. Deluzio, U.P. Wyss, B. Zee, P.A. Costigan, and C. Sorbie. Principal component models of knee kinematics and kinetics: Normal vs. pathological gait patterns. *Human Movement Science*, 16:201–217, 1997.

[34] J.S. Dufek, B.T. Bates, N. Stergiou, and C.R. James. Interactive effects between group and single-subject response patterns. *Hum Mov Sci*, 14(3):301–323, 1995.

[35] A. Duhamel, J.L. Bourriez, P. Devos, P. Krystkowiak, A. Destee, P. Derambure, and L. Defebvre. Statistical tools for clinical gait analysis. *Gait Posture*, 20:204–212, 2006.

[36] Y. Ehara, H. Fujimoto, S. Miyazaki, M. Mochimaru, S. Tanaka, S. Yamamoto, S. Barker, R. Craik, W. Freedman, N. Herrmann, and H. Hillstrom. Accuracy, reliability, and validity of a spatiotemporal gait analysis system. *Gait and Posture*, 5(3):251–255, 1997.

[37] P.H.C. Eilers and B.D. Marx. Flexible smoothing with β-splines and penalties. *Statistical Science*, 11(2):89–121, 1996.

[38] B. Elliott. Biomechanics: An integral part of sport science and sport medicine. *Journal of Science and Medicine in Sport*, 2(4):299–310, 1999.

[39] J. Fan and S.K. Lin. Test of significance when data are curves. *Journal of the American Statistical Association*, 93, 1998.

[40] J.J. Faraway and J. Hu. Modeling variability in reaching motions. In *Proceedings of the SAE Digital Human Modeling Conference*, 2001.

[41] R. Ferber, I. McClay Davis, D.S. Williams, and C. Laughton. A comparison of within- and between-day reliability of discrete 3d lower extremity variables in runners. *J Orthop Res*, 20(6):1139–1145, 2002.

[42] A. Forner-Cordero, O. Levin, Y. Li, and S.P. Swinnen. Principal component analysis of complex multijoint coordinative movements. *Biological Cybernetics*, 93:63–78, 2005.

[43] C. Glymour, D. Madigan, D. Pregibon, and P. Smyth. Statistical themes and lessons for data mining. *Data Mining and Knowledge Discovery*, 1(1):11–28, 1997.

[44] D. Graham and A.J. Harrison. An investigation of schema theory applied to the biomechanics of the sprint starts in athletics. In *Proceedings of the XXIV ISBS Symposium*, 2006.

[45] E. Growney, D. Meglan, M. Johnson, T. Cahalan, and K.N. An. Repeated measures of adult normal walking using a video tracking system. *Gait Posture*, 6(2):147–162, 1997.

[46] P. Hall, H.G. Mller, and J.L. Wang. Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, 34(3):1493–1517, 2006.

[47] J. Hamill, J. M. Haddad, and E.A. van Emmerik. Using coordination measures for movement analysis. In *Proceedings of the XXIII ISBS Symposium*, 2005.

[48] J. Hamill, J.M. Haddad, and E.A. van Emmerik. Overuse injuries in running: do complex analysis help our understanding? In *Proceedings of the XXIV ISBS Symposium*, 2006.

[49] J. Hamill, R.E. van Emmerik, B.C. Heiderscheit, and L. Li. A dynamical systems approach to lower extremity running injuries. *Clin Biomech*, 14(5):297–308, 1999.

[50] D. Hand, H. Mannila, and P. Smyth. Principles of data mining. *The MIT Press.*, 2001.

[51] A.J. Harrison, W. Ryan, and K. Hayes. Functional data analysis of joint coordination in the development of vertical jump performance. *Sports Biomechanics*, 6(2):199–214, 2007.

[52] H. Hatze. Motion variabilityits definition, quantification, and origin. *J Mot Behav*, 18(1):5–16, 1986.

[53] J.M. Hausdorff. Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking. *Human Movement Science*, 2007.

[54] J. Himberg. *From insights to innovations: data mining, visualization, and user interfaces.* PhD thesis, Helsinki University of Technology, 2004.

[55] A.L. Hof. Scaling gait data to body size [1]. *Gait Posture*, 4(3):222–223, 1996.

[56] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:498–520, 1933.

[57] J.L. Hudson. Sports biomechanics in the year 2000: Some observations on trajectory. In *Proceedings of XVIII International Symposium of Biomechanics in Sports*, 2000.

[58] International Association of Athletics Federations. *IAAF Competition Rules 2006-2007*, 2006.

[59] I.T. Joliffe. *Principal Component Analysis*. Springer, New York, NY.

[60] I.T. Joliffe and B.J.T. Morgan. Principal component analysis and exploratory factor analysis. *Statistical Methods in Medical Research*, 1, 1992.

[61] M.P. Kabada, H.K. Ramakrishnan, and M.E. Wootten. Measurement of lower extremity kinematics during level walking. *J Orthop Res*, 8(3):383–392, 1990.

[62] M.P. Kabada, H.K. Ramakrishnan, M.E. Wootten, J. Gainey, G. Gorton, and G.V. Cochran. Repestability of kinematic, kinetic, and electromyographic data in normal adult gait. *J Orthop Res*, 7(6):849–860, 1989.

[63] J. Kepner and R. Kim. Cluster detection in databases: The adaptive matched filter algorithm and implementation. *Data Mining and Knowledge Discovery*, 7(1):57–79, 2003.

[64] M. Kermit and O. Tomic. Independent component analysis applied on gas sensor array measurement data. *IEEE Sensors Journal*, 3(2):218–228, 2003.

[65] A. Kocsor and L. Toth. Kernel-based feature extraction with a speech technology application. *IEEE Transactions on Signal Processing*, 52:2250–2263, 2004.

[66] W. Kroll. Reliability theory and research decision in selection of a criterion score. *Research Quarterly*, 38, 1957.

[67] M.J. Kurz and N. Stergiou. Effect of normalization and phase angle calculations on continuous relative phase. *Journal of Biomechanics*, 35, 2002.

[68] M.J. Kurz and N. Stergiou. *Innovative analysis of human movement.* Human Kinetics, 2004.

[69] A. Lees. Technique analysis in sports: a critical review. *Journal of Sports Sciences*, 20:813–828, 2002.

[70] M.W. Lenhoff, T. Santner, J. Otis, M.G.E. Petreson, B.J. Williams, and S.I. Backus. Bootstrap prediction and confidence bands: a superior statistical method for analysis of gait data. *Gait Posture*, 9:10–17, 1997.

[71] S.E. Leurgrans, R.A. Moyeed, and B.W. Silverman. Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society - Series B*, 55(3):725–740, 1993.

[72] D.G. Liebermann, L. Katz, M.D. Hughes, R.M. Bartlett, J. Mcclements, and I.M. Franks. Advances in the application of information technology to sport performance. *Journal of Sports Sciences*, 20:755–769, 2002.

[73] M. Loewenstein and J. Spletzer. General and specific training. evidence and implications. *The Journal of Human Resources*, 34(4):710–733, 1999.

[74] E. Mller, U. Benko, C. Raschner, and H. Schwameder. Specific fitness training and testing in competitive sports. *Medicine Science in sports exercises*, 32(1):216–220, 2000.

[75] M. Mller. Decomposition of variability in the execution of goal oriented tasks: three components of skill improvement. *J. Exp. Psychol. Hum. Percept. Perform.*, 30(1):212–233, 2004.

[76] K.C. Moisio, D.R. Sumner, S. Shott, and D.E. Hurwitz. Normalization of joint moments during gait: A comparison of two techniques. *J Biomech*, 36(4):599–603, 2003.

[77] R.D. Mullineaux, R.M. Bartlett, and S. Bennett. Research design and statistics in biomechanics and motor control. *Journal of Sports Sciences*, 19:739–760, 2001.

[78] M.P. Murray, G.N. Guten, L.A. Mollinger, and G.M. Gardner. Kinematic and electromyographic patterns of olympic race walkers. *The American Journal of Sports Medicine*, 11(2):68–74, 1983.

[79] H.F. Neumann, J. Krug, and D. Gohlitz. Coordinative threshold in race walking. In *Proceedings of the ISBS Symposium*, 2006.

[80] B.M. Nigg and M. Bobbert. On the potential of various approaches in load analysis to reduce the frequency of sports injuries. *Journal of Biomechanics*, 23(1):3–12, 1990.

[81] B. Odegard, N. Bulat, B. Savilonis, and Y. Mendelson. An analysis of race walking styles using a 2-dimensional mathematical knee model. In *Bioengineering Conference, Proceedings of the IEEE 1997 23rd Northeast*, 1997.

[82] S.J. Olney, M.P. Griffin, and I.D. McBride. Multivariate examination of data from gait analysis of persons with stroke. *Phys Ther*, 78(8):814–828, 1998.

[83] D. Ormoneit, M.J. Black, T. Hastie, and H. Kjellstrm. Representing cyclic human motion using functional analysis. *Image and Vision Computing*, 23:1264–1276, 2005.

[84] D. Ormoneit, H. Sidenbladh, M.J. Black, T. Hastie, and D.J. Fleet. Learning and tracking human motion using functional analysis. In *IEEE Workshop on Human Modeling, Analysis and Synthesis*, 2000.

[85] F. O'Sullivan. Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association*, 81(393):96–103, 1986.

[86] A. Pascucci. *Toward a PhD Thesis on Pattern Recognition*. PhD thesis, Università degli Studi di Roma - La Sapienza, 2006.

[87] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572, 1901.

[88] A. Pedotti and C. Frigo. Quantitative analysis of locomotion for basic research and clinical applications. *Functional Neurology*, 7(4):47–56, 1992.

[89] B.T. Peters, M.H. Jeffrey, B.C. Heiderscheit, R.E.A. Van Emmerik, and J. Hamill. Limitations in the use and interpretation of continuous relative phase. *Journal of Biomechanics*, 36, 2003.

[90] E. Preatoni. *Innovative Methods for the Analysis of Sports Movements and for the Longitudinal Monitoring of Individual Motor Skills*. PhD thesis, Politecnico di Milano, 2007.

[91] E. Preatoni, A. La Torre, and R. Rodano. A biomechanical comparison between racewalking and normal walking stance phase. In *Proceedings of the ISBS Symposium*, 2006.

[92] R.M. Queen, M.T. Gross, and H. Liu. Repeatability of lower extremity kinetics and kinematics for standardized and self-selected running speeds. *Gait and Posture*, 23(3):282–287, 2006.

[93] I.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Number 22. Inc., New York, 1997.

[94] I.O. Ramsay and B.W. Silverman. *Applied functional Data Analysis*. Inc., New York, 2002.

[95] J.O. Ramsay and C.J. Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society Series B*, 53, 1991.

[96] J.O. Ramsay and X. Li. Curve registration. *Journal of the Royal Statistical Society Series B*, 60(2):351–363, 1998.

[97] R. Rodano. *Bioingegneria dello sport*, pages 513–529. Patron, 2003.

[98] T. Ruchlewicz, R. Staszkiewicz, W. Chwals, and J. Laska. Oscillations of the centre of gravity in world elite race walkers. *Physical Education and Sport*, 50, 2006.

[99] O.M. Rutherford and D.A. Jones. The role of learning and coordination in strength training. *European Applied Physiology*, 55, 1986.

[100] W. Ryan, A. J. Harrison, and K. Hayes. Functional data analysis in biomechanics: a case study of knee joint vertical jump kinemtaics. *Sports Biomechanics*, 5, 2006.

[101] H. Sadeghi. Local or global asymmetry in gait of people without impairments. *Gait and Posture*, 17:197–204, 2003.

[102] H. Sadeghi, P. Allard, F. Barbier, S. Sadeghi, S. Hinse, R. Perrault, and H. Labelle. Main functional roles of knee flexors/extensors in able-bodied gait using principal component analysis. *The Knee*, 9:47–53, 2002.

[103] H. Sadeghi, P. Allard, and M. Duhaime. Functional gait asymmetry in able-bodied subjects. *Human Movement Science*, 16:243–258, 1997.

[104] H. Sadeghi, F. Prince, S. Sadeghi, and H. Labelle. Principal component analysis of the power developed in the flexion/extension muscles of the hip in able-bodied gait. *Medical Engineering Physics*, 22:703–710, 2000.

[105] R.A. Schmidt and C.A. Wrisberg. *Motor learning and performance*. Human Kinetics, 2004.

[106] H. Schwameder. Issues and challenges of applied sport biomechanics research. *Journal of Biomechanics*, 40(2):S14, 2007.

[107] R. Shiavi and P. Griffin. Representing and clustering electromyographic gait patterns with multivariate techniques. *Med Biol Eng Computing*, pages 605–611, 1981.

[108] B.W. Silverman. Incorporating parametric effects into functional principal component analysis. *Journal of the Royal Statistical Society Series B*, 57(4):673–689, 1995.

[109] J.P. Stevens. *Applied Multivariate Statistics for the Social Sciences*. Lawrence Erlbaum Associates Publishers, 2002.

[110] E. Thelen and B.D. Ulrich. Hidden skills: a dynamic systems analysis of treadmill stepping during the first year. *Monographs of the Society for Research in Child Development*, 56(1):1–104, 1991.

[111] J.R. Thomas and J.K. Nelson. *Research methods in physical activity.* Human Kinetics, 1996.

[112] M. Tingley, C. Wilson, E. Biden, and W.R. Knight. An index to quantify normality of gait in young children. *Gait Posture*, 16, 2002.

[113] N.F. Troje. Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2(5):371–387, 2002.

[114] R.E.A. van Emmerik and R.C. Wagenaar. Effects of walking velocity on relative phase dynamics in the trunk in human walking. *Journal of Biomechanics*, 29(9):1175–1184, 1996.

[115] G.A. Westerfield. The use of biomechanics in the judging of race walking. In *Proceedings of the International Race Walk Forum*, 2007.

[116] D.A. Winter. Kinematic and kinetic patterns in human gait: variability and compensating effects. *Human Movement Science*, 3(1-2):51–76, 1984.

[117] S. Wolf, T. Loose, M. Schablowski, L. Doederlein, R. Rupp, H.J. Gerner, G. Bretthauer, and R. Mikut. Automated feature assessment in instrumented gait analysis. *Gait Posture*, 23:331–338, 2006.

[118] M.E. Wootten, M.P. Kadaba, and G.V.B. Cochran. Dynamic electromyography i: numerical representation using principal component analysis. *J Orthop Res*, pages 247–258, 1990.

[119] A.T. Wrigley, W.J. Albert, K.J. Deluzio, and J.M. Stevenson. Differentiating lifting technique between those who develop low back pain and those who do not. *Clinical Biomechanics*, 20:254–263, 2005.

[120] A.T. Wrigley, W.J. Albert, K.J. Deluzio, and J.M. Stevenson. Principal component analysis of lifting waveforms. *Clinical Biomechanics*, 21(6):567–578, 2006.

[121] J. Wu, J. Wang, and L. Liu. Feature extraction via kpca for classification of gait patterns. *Human Movement Science*, 2007.

[122] S. Yamamoto, Y. Suto, H. Kawamura, T. Hashizume, S. Kakurai, and S. Sugahara. Quantitative gait evaluation of hip diseases using principal component analysis. *Journal of Biomechanics*, 16:717–726, 1983.

[123] K.A. Yeomans and P.A. Golder. The guttman-kaiser criterion as a predictor of the number of common factors. *The Statistician*, 31(3):221–229, 1982.

[124] H. Yoon, K. Yang, and C. Shahabi. Feature subset selection and feature ranking for multivariate time series. *IEEE Transactions On Knowledge And Data Engineering*, 17(9):1186–1198, 2005.

[125] V.M. Zatsiorsky and V.N. Seluyanov. *The mass and inertia characteristics of the main segments of the human body*, pages 1152–1159. Human Kinetics, Champaign, IL, 1983.