A multivariate gait data analysis technique: application to knee osteoarthritis

J L Astephen* and **K J Deluzio**
School of Biomedical Engineering, Dalhousie University, Halifax, Canada

Abstract: Modern gait analysis is a powerful non-invasive tool for calculating the mechanical factors involved in pathological processes such as knee osteoarthritis (OA). Although very accurate measurements can be made, the clinical applicability and widespread use of gait analysis have been hindered by a lack of appropriate data analysis techniques for reducing and analysing the resulting large volumes of highly correlated gait data. This paper introduces a multidimensional gait data analysis technique that simultaneously considers multiple time-varying and discrete measures, exploiting the correlation structure between and within the measures. The multidimensional analysis technique was used to detect discriminatory mechanical features of knee OA gait patterns that involved interacting changes in several gait measures, at specific time portions of the gait cycle. The two most discriminatory features described a dynamic alignment difference and a loading response difference with knee OA.

Keywords: gait analysis, data analysis, principal component analysis, discriminant analysis, knee osteoarthritis, multivariate statistics

NOTATION

a_{DP}	knee angle in the distal–proximal direction
a_{LM}	knee angle in the lateral–medial direction
a_{PA}	knee angle in the posterior–anterior direction
$\boldsymbol{a}$	vector of linear discriminant function coefficients for features
BMI	body mass index
c_1	stride length
c_2	velocity
c_3	stance percentage
c_4	stance time
c_5	hip–knee–angle
c_6	standing knee flexion angle
c_7	medial joint space
c_8	body mass index
coeff	vector of linear discriminant function coefficients for original variables
f_{DP}	knee bone-on-bone force in the distal proximal direction
f_{LM}	knee bone-on-bone force in the lateral–medial direction
f_{PA}	knee bone-on-bone force in the posterior–anterior direction
HKA	hip–knee–ankle
m_{DP}	knee moment in the distal–proximal direction
m_{LM}	knee moment in the lateral–medial direction
m_{PA}	knee moment in the posterior–anterior direction
OA	osteoarthritis
PC	principal component
PCA	principal component analysis
$\mathbf{X}$	matrix of standardized original variables
y	linear discriminant function
$\bar{y}$	classification cut-off value between two groups of discriminant scores
$\mathbf{Z}$	matrix of principal component scores for each subject on retained features

Subscripts

DP	distal–proximal
LM	lateral–medial
PA	posterior–anterior

1 INTRODUCTION

Modern gait analysis is a powerful non-invasive tool that offers a unique means of measuring the biomechanical response to diseases of the musculoskeletal system, such as knee osteoarthritis (OA). The clinical applicability and widespread use of gait analysis, however, have been greatly hindered by a lack of appropriate gait data analy-

The MS was received on 6 August 2003 and was accepted after revision for publication on 30 April 2004.
** Corresponding author: School of Biomedical Engineering, Dalhousie University, 5981 University Avenue, Halifax, Nova Scotia, B3H 3J5 Canada.*

sis techniques for reducing and clinically interpreting large volumes of correlated gait data [**1**].

The large volume of data resulting from modern gait analysis is highly complex, multidimensional and correlated. Most gait data appear as temporal waveforms representing specific joint measures throughout the gait cycle (i.e. joint angles and moments). A common method of analysing gait data is to define and extract parameters from the waveforms (ranges, peak values, etc.). While this accomplishes data reduction through the subjective selection of data, the resulting set of parameters is highly correlated and the shape of the waveforms is not utilized. The problem of correlation has been addressed to some extent through the application of multivariate techniques such as principal component analysis (PCA) of gait parameters [**2**, **3**]. However, extracting predefined parameters from atypical waveforms is subjective, and parameters neglect the temporal information in gait waveforms.

In 1997, Deluzio *et al.* [**4**] recognized the strong correlations between the time samples of gait waveforms and introduced PCA, a multivariate statistical technique, to the analysis of kinematic and kinetic gait waveform measures. Using each percentage of the gait cycle as an individual variable in the analysis, PCA was used to extract representative waveforms, called features, which described the major modes of variation between subject waveforms. These workers were able to localize, within the gait cycle, deviations from normal gait. This waveform analysis technique retained the temporal information within the gait waveforms, but waveform measures were analysed individually, ignoring interrelationships between the waveform measures and the hierarchy of discriminatory power among the measures. Also, discrete measures, such as frontal plane knee alignment, are also important to the mechanics of knee OA and should be included in an analysis.

In 2001, Chau [**5**, **6**] produced a two-part critical review of current gait data analysis techniques. Previous gait analyses have utilized several data analysis techniques to recognize the multivariate nature of gait data, such as multiple correspondence analysis [**7**], fuzzy analysis [**8**, **9**] and wavelets [**10**]. However, no previous techniques have considered both the temporal correlation within gait waveform measures and the correlation structure between the gait measures. Chau expressed the need in gait analysis for more multivariate analysis techniques that quantified differences in gait waveforms and allowed for simultaneous interpretation of multiple gait signals.

In this paper, a multivariate gait data analysis technique is introduced that simultaneously analyses multiple waveform measures as well as discrete measures that are important to knee OA. Utilizing two multivariate statistical techniques, namely PCA and discriminant analysis, the analysis technique retained both the temporal correlation structure of waveform gait measures as well as the correlations between measures. Asymptomatic subjects and patients with severe knee OA were chosen for the discrimination procedure based on the ease of separation of these groups, in order to develop and illustrate the technique.

2 METHODS

2.1 Subjects

The data for this analysis consisted of an asymptomatic group and a group of patients with end-stage knee OA. The normal subject group consisted of 63 asymptomatic elderly volunteers. These subjects were over 45 years of age, were pain free and had no record of surgery to the lower limb and no evidence or history of arthritic disease at the time of testing. The OA patient group consisted of 50 elderly patients with end-stage knee OA, who were evaluated prior to total knee replacement surgery. Anthropometrics for all subjects are summarized in Table 1.

2.2 Gait analysis

Three-dimensional gait patterns of subjects were studied with a validated [**11**] optoelectronic gait analysis system [**12**] that incorporated a standardized radiographic technique [**13**]. Standardized radiographs for all subjects were taken with the knee in a natural standing position and included an anterio-posterior view of the hip and knee as well as a lateral knee view for the test leg only. X-rays were used to measure static knee alignment, geometry and muscle moment arms and locations of bone landmarks. This allowed for accurate transformations from surface marker locations to joint centres. Three-dimensional locations of six infrared-light-emitting diodes were measured at 50 Hz. A force plate

Table 1 Subject anthropometrics

Subject group	N	Age* (years)	Height†‡ (m)	Body mass†‡ (kg)	Body mass index‡
Normal subjects	63	46–80; average, 65	1.68 (0.09)	80.68 (15.4)	29.32
Knee OA patients	50	53–84; average, 70	1.63 (0.09)	86.63 (22.1)	28.81

* The ages of the subjects are given in the form of a range and an average.
† The average heights and masses are given, as well as their standard deviations.
‡ No significant difference.

was synchronized with the global coordinate system of the camera with a motion calibration frame.

An inverse dynamics procedure was used to calculate three-dimensional knee joint angles, moments and forces with respect to the tibia plateau. The three-dimensional sign convention for the angles, moments and forces follows an anatomically based coordinate system. Knee angles were defined according to Grood and Suntay [**14**] and moments were expressed as knee moments, normalized to a subject's body mass [**12**].

Scaled radiographic measurements helped to construct a subject specific knee model used to estimate the forces that generated the knee joint moments. A simple knee model was used to calculate the joint contact forces at the knee in the posterior–anterior and distal–proximal directions. The knee was modelled as a two-dimensional structure that could be positioned in three-dimensional space [**15**]. In this model, the sagittal plane moment and the quadriceps or hamstring moment arm were used to estimate the muscle force required to generate the net knee flexion moment. In the model, knee extension was produced solely by the quadriceps muscles, and knee flexion was produced solely by the hamstring muscles. Contributions from other muscles (such as the gastrocnemius muscles) and soft tissues were ignored by the model. It was also assumed that there was no co-contraction of the quadriceps and hamstring muscles and muscle moment arms were considered constant. This major assumption of no co-contraction underestimates the magnitude of the joint contact forces at the knee. The muscle force was computed by dividing the net moment by the appropriate moment arm and estimating the orientation of the muscle force vector. Segmental inertial properties were estimated using regression equations based on subject specific anthropometrics [**16**].

2.3 The data

Both time varying and constant measures were included in the analysis. Nine waveform measures and eight discrete measures were simultaneously analysed (Table 2). *Waveform measures* are dynamic gait measurements that vary continuously in time during gait. Waveform measures are displayed graphically as curves; each curve is time normalized to represent one complete gait cycle, defined as the time from initial foot contact with the ground until second foot contact with the same foot. A waveform measure is therefore defined by 101 gait cycle values, one for each percentage of the gait cycle. *Discrete measures* are specified by a single value and include both stride characteristics and constant parameters that are related to knee OA. A complete list of the measures included in the analysis is given in Table 2. With nine waveform measures, each specified by 101 data points, and eight discrete measures, the original data included 917 gait variables.

2.4 Statistical methods: the multidimensional gait data analysis technique

The multidimensional gait data analysis technique was divided into three steps: feature extraction, discrimination and interpretation. The feature extraction step used the multivariate statistical technique, PCA [**17**], to extract multidimensional gait features that capture the greatest variation within the gait data. The second step, discrimination, utilized a linear discriminant analysis to determine the hierarchy among the extracted features in terms of their relative power in separating the normal and OA gait patterns [**18**].

Interrogation of the underlying linear structure of the model provided a biomechanical interpretation of the discriminatory features. Each discriminatory feature was interpreted in terms of firstly, the relative contribution of each percentage of the gait cycle and, secondly, the relative contribution of each original gait measure. For all major contributing gait measures, the direction of change necessary to classify a gait pattern as normal or abnormal was determined.

2.5 Feature extraction

The original 917 gait variables were standardized to have a zero mean and unit variance. The 113 subject obser-

Table 2 The 17 input measures considered in the multidimensional analysis

Analysis input measures					
Waveform measures			Discrete measures		
Symbol	Description	Units	Symbol	Description	Units
a_{PA}	Knee abduction–adduction angle	deg	c_1	Stride length	m
a_{LM}	Knee flexion–extension angle	deg	c_2	Velocity	m/s
a_{DP}	Knee internal–external rotation angle	deg	c_3	Stance percentage	
m_{PA}	Knee abduction–adduction moment normalized to body mass	N m/kg	c_4	Stance time	s
m_{LM}	Knee flexion–extension moment normalized to body mass	N m/kg	c_5	Hip–knee–ankle angle	deg
m_{DP}	Knee internal–external rotation moment normalized to body mass	N m/kg	c_6	Standing knee flexion angle	deg
f_{PA}	Posterior–anterior knee force normalized to body mass	N/kg	c_7	Medial joint space	mm
f_{LM}	Lateral–medial knee force normalized to body mass	N/kg	c_8	Body mass index	
f_{DP}	Distal–proximal knee force normalized to body mass	N/kg			

vations on the standardized original variables were contained in the (113×917) analysis matrix $\mathbf{X}$, given by

$$\mathbf{X} = [X_{a_{\mathrm{PA}}}\, X_{a_{\mathrm{LM}}}\, X_{a_{\mathrm{DP}}}\, X_{m_{\mathrm{PA}}}\, X_{m_{\mathrm{LM}}}\, X_{m_{\mathrm{DP}}}\, X_{f_{\mathrm{PA}}}\, X_{f_{\mathrm{LM}}}\, X_{f_{\mathrm{DP}}}\, X_d] \tag{1}$$

$\mathbf{X}$ is partitioned into ten matrices. The first nine partitions represent the waveform measures; $\mathbf{X}_d$ contains the eight discrete measures. The subscripts and ordering of the waveform measures are defined in Table 2.

The major features of variation in the original data, contained in the matrix $\mathbf{X}$, were extracted with the multivariate statistical technique, PCA. PCA orthogonally rotates the original 917 standardized variables into 917 new uncorrelated principal components (PCs), called features, which maximally explain the variability in the original data [**19**].

Mathematically, PC features are the eigenvectors of the covariance matrix of $\mathbf{X}$, with variances equal to the magnitudes of their corresponding eigenvalues. Ordered according to the size of their eigenvalues, the features describe decreasing amounts of data variation. The majority of data variation can usually be described by a small subset of these features. The data can therefore be accurately described by a smaller number of features than the original variables. A 90 per cent trace criterion was used to determine the number of features to retain in the analysis [**19**].

The data for all subjects was projected on to the subspace spanned by the reduced set of features that explain at least 90 per cent of the variation. The orthogonal projection of an observation on to a particular feature is called a PC score. The information contained in the original data was more compactly described by a reduced set of PC scores, contained in the matrix $\mathbf{Z}$ given by

$$\mathbf{Z} = \mathbf{X} \times \mathbf{U}_r = \mathbf{X} \times [\boldsymbol{u}_1 \dots \boldsymbol{u}_r] = [z_1 \dots z_r] \tag{2}$$

where:

$\mathbf{Z} = (113 \times r)$ matrix of PC scores for each subject on all retained features
r = number of retained features ($r \ll 917$)
$\mathbf{X} = (113 \times 917)$ matrix of original standardized variables
$\mathbf{U}_r = (917 \times r)$ matrix of the first r eigenvectors of the covariance matrix of $\mathbf{X}$
$\boldsymbol{u}_i$ = vector containing the ith eigenvector (feature i)
$z_i = (113 \times 1)$ vector of the PC score observations on feature i, for each subject

2.6 Discrimination

A linear discriminant analysis was applied to the features that were extracted from the original data with the PCA and retained during the feature extraction phase of the analysis. A backward elimination stepwise discrimination procedure was employed to determine the subset of retained features that optimally separated the two groups of PC scores. Features that did not significantly contribute to the discriminatory power of the model, measured by the likelihood ratio criterion Wilk's lambda, were successively removed from the model until the most discriminatory model was achieved [**18**].

A discriminant function y was defined as a linear combination of the subset of discriminatory features from the stepwise procedure, using Fisher's approach to linear discrimination [**18**]. The discriminant function defined the optimal boundary of separation between the normal and OA groups of PC scores. The discriminant function was used as a group classification rule and an indicator of multidimensional group separation [**20**]. The magnitude of a feature's coefficient in the discriminant function quantified the relative importance of that feature in the multivariate separation of the two groups [**20**]. Discriminant function coefficients were compared to establish a hierarchy of discriminatory power among the features.

A discriminant score was calculated for each subject by applying the subject's set of PC scores to the discriminant function formula. The discriminant score was the optimal univariate indicator of gait pattern differentiation; differences in subject discriminant scores optimally reflected multidimensional gait pattern differences in a single value. The discriminant score represented a disease severity and classification index. The magnitude of a subject's discriminant score quantified the level of abnormality of OA gait. A smaller discriminant score was associated with a more abnormal gait pattern (i.e. further from the group of normal gait patterns); a larger discriminant score indicated a more normal gait pattern. A cut-off value for group classification, $\bar{y}$, was defined as the midpoint between the mean normal discriminant score and the mean OA discriminant score. A discriminant score greater than $\bar{y}$ classified a subject as normal; a score less than $\bar{y}$ classified a subject as OA.

The same data set was used to derive the discrimination model and to classify subjects. Therefore, an unbiased cross-validation misclassification error rate was used to validate the ability of the discriminant model in classifying the subjects [**18**].

2.7 Interpretation

Discriminatory features were interpreted in terms of, firstly, the relative contribution of each percentage of the gait cycle and, secondly, the relative contribution of each original gait measure. Each feature was defined as a linear combination of the original 917 variables, with coefficients that corresponded to the discrete measures and to every percentage of the gait cycle for each of the nine waveform measures. The relative importance of an original variable within a feature was quantified by the correlation between the variable and the feature. This quantity is referred to as percentage variation explained because it represents the amount of variation in the original variable explained by the particular feature

$$\text{Percentage variation explained } (i,j) = \frac{100u_{ij}\sqrt{\text{var}(z_i)}}{\text{var}(\boldsymbol{x}_j)} \tag{3}$$

where:

Percentage variation explained (i,j) = percentage variation explained of the original variable $\boldsymbol{x}_j$
u_{ij} = coefficient of $\boldsymbol{x}_j$ in feature i
z_i = vector of PC scores for all subjects, corresponding to feature i
$\text{var}(z_i)$ = variance of feature i
$\boldsymbol{x}_j$ = vector containing all observations on the j^{th} original standardized variable
$\text{var}(\boldsymbol{x}_j)$ = variance of $\boldsymbol{x}_j$

Hierarchies of the relative contributions of the measures to the features were established by comparing the values of the percentage variation explained for the 17 input measures for each feature. An average value of the percentage variation explained over the gait cycle was used to represent the contribution of a waveform measure, where these values represented the overall correlation of the feature to the waveform measures and allowed them to be compared with the discrete measures:

$$\text{Percentage variation explained } (w_j, i) = \frac{\sum_m^{101} 100u_{iw_{jm}}\sqrt{var\ (z_i)}/\text{var}\,(\boldsymbol{x}_{w_{jm}})}{101} \tag{4}$$

where

Percentage variation explained (w_j, i) = average value of percentage variation explained over all gait cycle values of the waveform measure j
$u_{iw_{jm}}$ = coefficient of feature i corresponding to the mth percentage of the gait cycle in the jth waveform measure; $m = 1, \ldots, 101$
z_i = vector of PC scores for all subjects, corresponding to feature i
$\text{var}\,(z_i)$ = variance of feature i
$\boldsymbol{x}_{w_{jm}}$ = vector of observations on the standardized variable corresponding to the mth gait cycle value of the jth waveform measure
$\text{var}\,(\boldsymbol{x}_{w_{jm}})$ = variance of $\boldsymbol{x}_{w_{jm}}$

To quantify the overall relative importance of each percentage of the gait cycle to a feature, an average value of percentage variation explained was calculated over the nine waveform measures, at each of the 101 gait cycle values:

$$\text{Percentage variation explained } (gc_m, i) = \frac{\sum_j^{9} 100u_{iw_{jm}}\sqrt{var\ (z_i)}/\text{var}\,(\boldsymbol{x}_{w_{jm}})}{9} \tag{5}$$

where

Percentage variation explained (gc_m, i) = average value of percentage variation explained over all nine waveform measures for the mth gait cycle value of feature i; $j = 1, \ldots, 9$, $k = 1, \ldots, 101$
$u_{iw_{jm}}$ = coefficient of the mth gait cycle value of the jth waveform measure in feature i
z_i = vector of PC scores for all subjects, corresponding to feature i
$\text{var}\,(z_i)$ = variance of feature i
$\boldsymbol{x}_{w_{jm}}$ = vector of observations on the standardized variable corresponding to the mth gait cycle value of the jth waveform measure
$\text{var}\,(\boldsymbol{x}_{w_{jm}})$ = variance of $\boldsymbol{x}_{w_{jm}}$

Examining the quantity given in equation (5) graphically over the gait cycle allowed a feature to be isolated to a particular gait event.

The major contributing measures of the features contributed to group separation in one of two ways. Either a larger or a smaller observation of the variable contributed to a more abnormal gait pattern. By expressing the linear discriminant function y equivalently as a linear combination of the original variables, the direction of change in an original variable associated with either normal or OA classification was determined:

$$y = \boldsymbol{a}' \cdot \boldsymbol{z} = \boldsymbol{a}' \cdot (\mathbf{U}_r' \boldsymbol{x}) = (\boldsymbol{coeff}) \cdot \boldsymbol{x} \tag{6}$$

where

y = linear discriminant function (i.e. a discriminant score for a particular subject)
$\boldsymbol{a}'$ = vector of the linear discriminant function coefficients
$\boldsymbol{z} = (r \times 1)$ vector of a subject's PC score observations for all retained k features
$\boldsymbol{x} = (917 \times 1)$ vector of a subject's observation on all original variables
$\mathbf{U}_m = (917 \times r)$ matrix of the first k eigenvectors of the covariance matrix of $\mathbf{X}$
$\boldsymbol{coeff} = (1 \times 917)$ vector of the discriminant function coefficients of the original variables

The polarity of the coefficient of an original variable in the discriminant function indicated the direction of change in that variable associated with normal or OA classification. A large positive term in the discriminant function contributed to a larger discriminant score, and therefore to normal group classification. Smaller positive or larger negative terms produced OA tendencies.

3 RESULTS

3.1 Feature extraction

Twenty-five features were extracted with PCA, cumulatively explaining 93 per cent of the original data variation. The percentage of the total variation explained by each of the 25 features ranged from 18.46 per cent for the first feature extracted to 0.58 per cent for the 25th feature extracted (Table 3). A backward stepwise discrimination procedure indicated that a 12-dimensional subset of these features was optimal for multivariate separation of the normal and OA groups. The 12 discriminatory features cumulatively explained 70 per cent of the variation in the data.

3.2 Discrimination

Multivariate F tests (Wilk's lambda, Pillai's trace and the, Hotelling-Lawley trace) confirmed significant mean differences between the groups of PC scores ($p<0.0001$). A linear discriminant function was defined that successfully separated the normal and OA groups, with a cross-validation misclassification error rate of 6% (Fig. 1).

The hierarchy of discriminatory importance of the 12 features was established by ordering the features according to the magnitudes of their coefficients in the discriminant function (Fig. 2). The first feature extracted with PCA, feature 1, was the most important feature to group separation and explained 18.5 per cent of the original data variation. Feature 20 was the second most discriminatory feature in the analysis. It explained 0.85 per cent of the original data variation (Table 3). Feature 20 was a relatively low-variance but very discriminatory feature.

3.3 Interpretation

To illustrate the interpretation phase of the analysis, the biomechanical interpretation of the two most discriminatory features will be described. Feature 1, the most discriminatory feature, had relatively equal contributions from seven major input measures, including the lateral–medial force f_{LM}, the adduction moment m_{PA}, the stance time c_4, the internal rotation moment m_{DP}, the hip–knee–ankle (HKA) angle c_5, the velocity c_2 and the standing flexion angle c_6 (Fig. 3). The values of percentage variation explained for these seven measures were at least 50 per cent of the percentage variation explained of the greatest contributor, the lateral–medial force.

Table 3 Summary of PCA and discriminant analysis

PC	Percentage variation explained	Rank in discrimination hierarchy
1	18.46	1
2	12.80	12
3	10.54	8
4	7.54	9
5	6.77	11
6	5.51	Insignificant
7	3.89	Insignificant
8	3.16	5
9	2.94	Insignificant
10	2.79	Insignificant
11	2.23	4
12	1.99	7
13	1.84	10
14	1.79	6
15	1.47	Insignificant
16	1.36	3
17	1.26	Insignificant
18	1.11	Insignificant
19	1.07	Insignificant
20	0.84	2
21	0.79	Insignificant
22	0.74	Insignificant
23	0.68	Insignificant
24	0.60	Insignificant
25	0.58	Insignificant

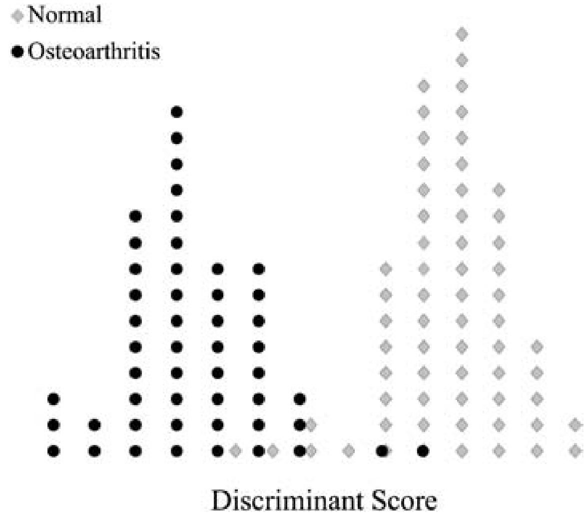

Fig. 1 Discriminant scores. The discriminant score histograms of 63 normal subjects (◆) and 50 knee OA subjects (●) are shown. The two groups of discriminant scores were well separated, with a cross-validation misclassification of 6 per cent

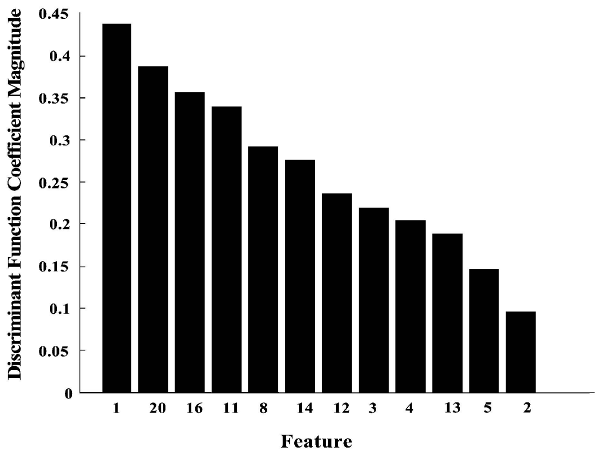

Fig. 2 Discrimination hierarchy. The magnitudes of the coefficients of the 12 discriminatory features are shown. Features 1 and 20 had the most discriminatory power

Feature 1 was important during the stance phase of the gait cycle, from approximately 20 to 60 per cent of the gait cycle (Fig. 4). The directions of change in the major contributing measures to feature 1 are summarized in Table 4. For instance, the coefficients of the adduction moment waveform at each percentage of the

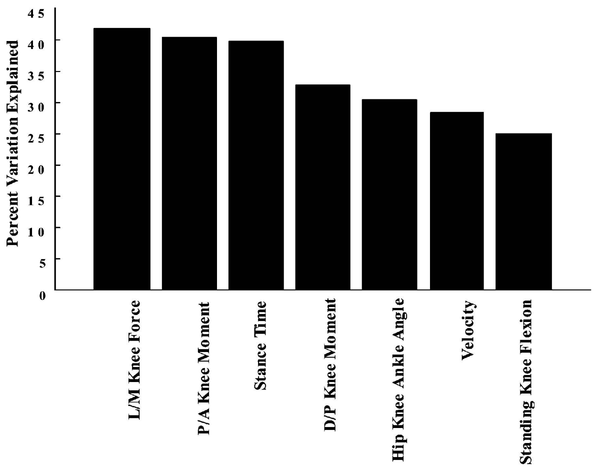

Fig. 3 Major contributing measures to feature 1. The values of the percentage variation explained for the seven major contributors to feature 1 are shown. Each major contributor had a percentage variation explained of at least 50 per cent of the maximum percentage variation explained

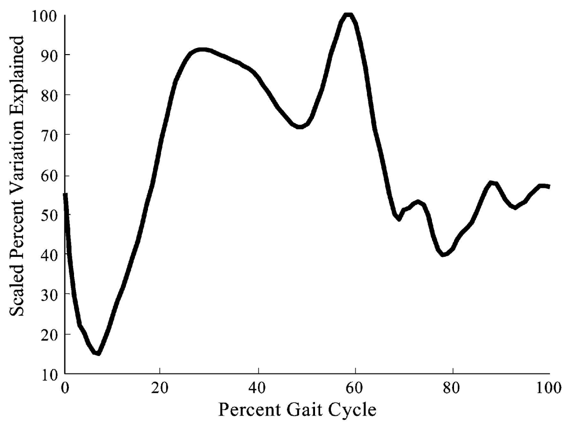

Fig. 4 Gait cycle importance to feature 1. The average value of the percentage variation explained for each per cent of the gait, over the nine waveform gait measures, is shown for feature 1. Feature 1 described a difference during the portion of the stance phase of the gait cycle from approximately 20 to 60 per cent, where the values of the percentage variation explained are highest

Table 4 Biomechanical interpretation summary of feature 1

Major contributor	Portion of gait cycle	Normal group	OA group
Waveform measures			
Lateral–medial knee force	20–60%	Larger	Smaller
Knee adduction moment	20–60%	Smaller	Larger
Knee internal rotation moment	20–40%	Smaller	Larger
	40–60%	Larger	Smaller
Discrete measures			
Stance time		Smaller	Larger
HKA angle		Larger	Smaller
Velocity		Larger	Smaller
Standing flexion angle		Larger	Smaller

gait cycle during stance were negative. Larger adduction moments tended to produce smaller discriminant scores, or more abnormal gait patterns. Larger adduction moments during stance were therefore a characteristic of the OA gait patterns.

Feature 20, the second most discriminatory feature, completely isolated the loading response phase of the gait cycle. This is the initial portion of the gait cycle immediately following heel strike (Fig. 5). The body mass index (BMI) c_8, a relative obesity measure, was the major contributing measure to feature 20. Its contribution to feature 20 was approximately seven times greater than any other measure. The coefficient of BMI in the discriminant function was negative. Larger BMI values produced tendencies towards OA group classification; smaller BMI values produced normal classification tendencies. Obesity was therefore an important contributor to abnormal gait patterns during the loading response phase of the gait cycle.

Although the BMI was the greatest contributing measure to feature 20, a univariate Student's t test on the mean BMI values for the groups indicated no significant difference between the normal BMI values and the OA BMI values (Table 1). The BMI is not a discriminatory parameter on its own. This would indicate that the BMI is an important characteristic of OA gait patterns when it exists in combination with the other observed differences during loading response.

4 DISCUSSION

Gait is a complex multifactorial process that produces large volumes of correlated data. An appropriate gait data analysis technique should retain and exploit the

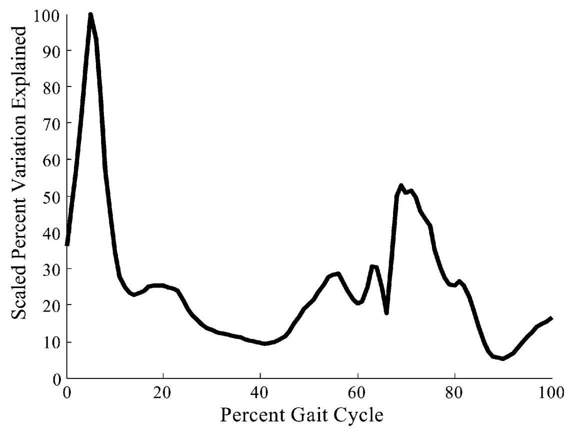

Fig. 5 Gait cycle importance to feature 20. The average values of the percentage variation explained for each per cent of the gait, over the nine waveform measures, are shown. Feature 20 is an early-stance loading response feature because the plot of the percentage variation explained peaks from approximately 0 to 12 per cent of the gait cycle

interrelationships between gait measures. In this study, a gait data analysis technique was introduced that successfully reduced multivariate gait data into biomechanically interpretable features. The features represented a multifactorial gait phenomenon that involved interacting changes in several measures, at specific portions of the gait cycle.

A single-valued discriminant index was defined that quantified the severity of knee OA gait patterns. The index was the optimal single quantity that represented the multiple differences in normal and knee OA gait patterns. Previous multivariate gait analyses have used binary measures [**4**] to provide overall gait assessments, but these methods ignore the hierarchy of discriminatory importance among the features. The OA index defined in this study represents a quantitative measure of disease severity that has the potential to be used as a diagnostic and treatment evaluation tool, or as an objective measure of outcome following a surgical procedure such as knee arthroplasty.

The most discriminatory feature extracted in this analysis, feature 1, represented a dynamic alignment difference with knee OA. The difference described involved simultaneous changes in several factors during the stance phase of the gait cycle, including the dynamic knee adduction moment, and the static frontal alignment angle, measured by the HKA angle. Several previous knee OA gait studies have identified the contribution of varus malalignment (HKA angle) [**21**, **22**] and the dynamic knee adduction moment [**23**, **24**] to knee OA gait patterns. There is controversy over the relationship between static malalignment, as measured by the HKA angle, and the knee adduction moment. Positive correlations between the HKA angle and the peak knee adduction moment have been identified in previous analyses [**21**, **23**, **25**, **26**]. However, Prodromos *et al.* [**27**] found no significant correlation between the HKA angle and the peak adduction moment. Inconsistency in the results may be because the analyses did not consider the dynamic nature of the adduction moment and the multivariate correlations between many gait measures, not only the HKA angle and the adduction moment. The discriminatory alignment feature that was detected in this analysis emphasized the need for multivariate analysis techniques in gait analysis. The results of this study identified not only the importance of alignment related measures to knee OA, but also how and when they interact during gait to produce the differences. Alignment-related differences contributed to abnormal gait patterns when they existed simultaneously during the stance phase of the gait cycle.

Detection of the second most discriminatory feature, feature 20, by the multivariate analysis technique strongly supports the importance and potential benefit of maintaining the time-dependent structure of gait waveforms in an analysis. Feature 20 completely isolated the portion of the gait cycle known as the loading response phase. Loading response represents the most demanding task during gait immediately following heel strike, when a subject's body mass is shifted from two supporting limbs to a single-limb stance. This task requires a great deal of coordination, shock absorbency and limb stability. This gait event has been previously hypothesized to be important to the disease process of knee OA [**28**], and this study quantitatively identified its importance in severe knee OA gait.

Feature 20 explained only a small portion of the variability in the original data (0.86 per cent), and yet it described a very important gait pattern difference between the two groups. The statistical literature has recognized the potential discriminatory power of small variance features extracted with PCA [**29**, **30**]. The orthogonality of the features extracted with PCs allows very subtle differences between groups to be detected.

The interpretation phase of the analysis is specific to the particular feature of interest and therefore more involved than the other phases of the analysis. Description of the biomechanical interpretation of the two most discriminatory features extracted with PCA was chosen because these features illustrate the two most important differences between the gait patterns of the normal and knee OA groups. Similar interpretation analyses could be applied to the remaining ten discriminatory features to describe more differences between the gait patterns.

While the multidimensional gait data analysis technique described in this paper was successful in comparing the gait patterns of OA patients with those of normal subjects, the two groups studied were from either end of the OA disease severity spectrum. Differences between the groups' gait patterns were expected and interpretable. It is not clear from this study that the same technique would be useful for discriminating patients with different pathologies or patients with more than one pathology. However, the technique was developed in such a way that it may be generalized to apply to other multivariate time-dependent data applications where differences may be more subtle.

5 CONCLUSIONS

Biomechanical changes that occur with knee OA involve multiple interacting factors that progress throughout the course of the disease. Exploiting the interrelationships between and within measures in the analysis of gait data may lead to a fuller understanding of the pathological mechanisms that may be responsible for disease initiation and/or progression. Extracting important and clinically useful information from gait data is a huge problem faced by clinicians. Results are commonly interpreted subjectively from a large number of highly correlated, time-varying and constant variables. The multidimensional technique described in this paper represented an objective and robust method of simultaneously reducing and analysing many interrelated time-varying and constant gait measures. The development of the technique was not specific to this particular application or even to gait analyses in general. It has the potential for use in a wide variety of multivariate data applications.

ACKNOWLEDGEMENTS

This study was supported by the Natural Sciences and Engineering Research Council of Canada. The authors would also like to thank Dr Urs Wyss, Dr Pat Costigan and Dr Charles Sorbie of the Human Mobility Research Center at Queen's University, Kingston, Canada, for their help and contribution in the collection of the motion analysis and radiography data used in this study.

REFERENCES

1 **Brand, R. A.** Assessing gait for clinical decisions. In Proceedings of the European Society of Biomechanics Conference, Rome, Italy, 1992, pp. 256–259.
2 **Sadeghi, H., Allard, P.** and **Duhaime, M.** Functional gait asymmetry in able bodied subjects. *J. Human Movement Sci.*, 1997, **16**, 243–258.
3 **Olney, S. J., Griffin, M. P.** and **McBride, I. D.** Multivariate examination of data from gait analysis of persons with stroke. *Phys. Therapy*, 1998, **78**, 814–828.
4 **Deluzio, K. J., Wyss, U. P., Zee, B., Costigan, P. A.** and **Sorbie, C.** Principal component models of knee kinematics and kinetics: normal vs. pathological gait patterns. *J. Human Movement Sci.*, 1997, **16**, 201–217.
5 **Chau, T.** A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods. *Gait Posture*, 2001, **13**(1), 49–66.
6 **Chau, T.** A review of analytical techniques for gait data. Part 2: neural network and wavelet methods. *Gait Posture*, 2001, **13**(2), 102–120.
7 **Loslever, P.** Error and data coding in the multi-dimensional analysis of human movement signals. *J. Engng Medicine*, 1993, **207**, 103–110.
8 **O'Malley, M. J., Abel, M. F., Damiano, D. L.** and **Vaughan, C. L.** Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters. *IEEE Trans. Rehabilitation Engng*, 1997, **5**(4), 300–309.
9 **Tan, T., Guan, L.** and **Burne, J.** A real-time image analysis system for computer-assisted diagnosis of neurological disorders. *Real-time Imaging*, 1999, **5**(4), 253–269.
10 **Wachowiak, M., Rash, G., Desoky, A.** and **Quesada, P.** Wavelet transforms for smoothing kinesiological data. In Proceedings of the 21st Annual Meeting of the American Society of Biomechanics, Clemson University, Clemson, South Carolina, 1997, p. 73.
11 **Deluzio, K. J., Wyss, U. P., Li, J.** and **Costigan, P. A.** A procedure to validate three-dimensional motion assessment systems. *J. Biomechanics*, 1992, **26**, 753–759.
12 **Costigan, P. A., Wyss, U. P., Deluzio, K. J.** and **Li, J.** Semiautomatic three-dimensional knee motion assessment system. *Med. Biol. Engng Computing*, 1992, **30**(3), 343–350.
13 **Siu, D., Cooke, T. D. V., Broekhoven, L. D., Lam, M., Fisher, B., Saunders, G.** and **Challis, T. W. A.** Standardized technique for lower limb radiography: practice, applications, and error analysis. *Investigative Radiogr.*, 1991, **26**, 71–77.
14 **Grood, E. S.** and **Suntay, W. J.** A joint coordinate system for the clinical description of three-dimensional motions: application to the knee. *J. Biomech. Engng*, 1983, **105**, 136–144.
15 **Li, J.** An integrated gait analysis system (QGAIT) for evaluation of individual loading patterns at knee joint during gait. Thesis/Dissertation, Queen's University, Kingston, Canada, 1992.
16 **Clauser, E. C., McConville, J. T.** and **Young, J. W.** Weight volume and centre of mass of segments of the human body. AMRL Technical Report, Wright–Patterson Air Force Base, Ohio (National Technical Information Service No. AD-710-622), 1969, pp. 69–70.
17 **Pearson, K.** On lines and planes of closest fit to systems of points in space. *Phil. Mag. B*, 1901, **2**, 559–572.
18 **Lachenbruch, P. A.** *Discriminant Analysis*, 1975 (Hafner Press, New York).
19 **Jackson, J. E.** *A User's Guide to Principal Components*, 1991 (John Wiley, New York).
20 **Johnson, R. A.** and **Wichern, D. W.** *Applied Multivariate Statistical Analysis*, 4th edition, 1998 (Prentice Hall, Englewood Cliffs, New Jersey).
21 **Olney, S. J., Griffin, M. P.** and **McBride, I. D.** Temporal, kinematic, and kinetic variables related to gait speed in subjects with hemiplegia: a regression approach. *Phys. Therapy*, 1994, **74**(9), 872–885.
22 **Sharma, L., Lou, C., Cahue, S.** and **Dunlop, D. D.** The mechanism of the effect of obesity in knee osteoarthritis: the mediating role of malalignment. *Arthritis Rheum.*, 2000, **43**(3), 568–575.
23 **Hurwitz, D. E., Ryals, A. B., Case, J. P., Block, J. A.** and **Andriacchi, T. P.** The knee adduction moment during gait in subjects with knee osteoarthritis is more closely correlated with static alignment than radiographic disease severity, toe out angle and pain. *J. Orthop. Res.*, 2002, **20**(1), 101–107.
24 **Kaufman, K. R., Hughes, C., Morrey, B. F., Morrey, M.** and **An, K.** Gait characteristics of patients with knee osteoarthritis. *J. Biomechanics*, 2001, **34**(7), 907–915.
25 **Andrews, M., Noyes, F. R., Hewett, T. E.** and **Andriacchi, T. P.** Lower limb alignment and foot angle are related to stance phase knee adduction in normal subjects: a critical analysis of the reliability of gait analysis data. *J. Orthop. Res.*, 1996, **14**, 289–295.
26 **Hilding, M. B., Lanshammar, H.** and **Ryd, L.** A relationship between dynamic and static assessments of knee joint load. Gait analysis and radiography before and after knee replacement in 45 patients. *Acta Orthop. Scand.*, 1995, **66**, 317–320.
27 **Prodromos, C. C., Andriacchi, T. P.** and **Galante, J. O.** A relationship between knee joint loads and clinical changes following high tibial osteotomy. *J. Bone Jt Surg.*, 1985, **67A**, 1188–1194.
28 **Radin, E. L., Yang, K. H., Riegger, C., Kish, V. L.** and **O'Connor, J. J.** Relationship between lower limb dynamics and knee joint pain. *J. Orthop. Res.*, 1991, **9**(3), 398–405.
29 **Jolliffe, I. T.** A note on the use of principal components in regression. *Appl. Statists*, 1982, **31**, 300–303.
30 **Hawkins, D. M.** and **Fatti, L. P.** Exploring multivariate data using the minor principal components. *Statistician*, 1984, **33**, 325–338.