# Knee Kinematic Signals Clustering for the Identification of Sagittal and Transverse Gait Patterns

N. Mezghani [1,2], M. Toumi [2], A. Fuentes [4],
A. Mitiche [3], N. Hagemeister, and J.A. de Guise [2]

[1] Centre de recherche du LICEF, TELUQ, Montreal (Quebec), Canada
[2] Laboratoire de recherche en imagerie et orthopédie (LIO) - Centre de recherche du
CHUM, Montreal (Quebec), Canada.
neila.mezghani@teluq.ca
[3] INRS- Énergie, matériaux et télécommunications, Montreal (Quebec), Canada
[4] Centre du genou EMOVI, Laval (Quebec), Canada

## Abstract

The purpose of this study is to investigate knee kinematic signals clustering by principal component analysis. The aim is to identify meaningful patterns in normal gait knee flexion/extension and tibial internal/external rotation signals. To preserve all of the information contained in these kinematics signals, the analysis uses the entire angle curve over a gait cycle rather then a few features extracted from this curve as done traditionally. To reduce processing complexity, the data dimensionality is reduced without loss of relevant information by projecting the gait curve onto a subspace of significant principal components (PCs). Gait patterns are then extracted by a discriminant analysis of the set of training data based on the PCs sign. The analysis identified two representation patterns for each of the flexion/extension (sagital plane) and the tibial internal/external rotation (transverse plane). These patterns were validated both by the clustering silhouette width and clinical interpretation.

**Keywords:** Gait pattern identification, knee kinematics, clustering, principal component analysis.

## 1 INTRODUCTION

The analysis of knee kinematic data during gait can inform about an individual's locomotion function. A characterization of normal gait, also called asymptomatic gait, by a few patterns representative of clearly distinct gait categories is necessary to understand the normal locomotion function [15]. Such a characterization faces two basic difficulties. First, there is a significant variability in the kinematic data of normal gait. Second, the data are given in the form of a vector of measurements of high dimension for each subject, causing their analysis to suffer from the curse of dimensionality [5].
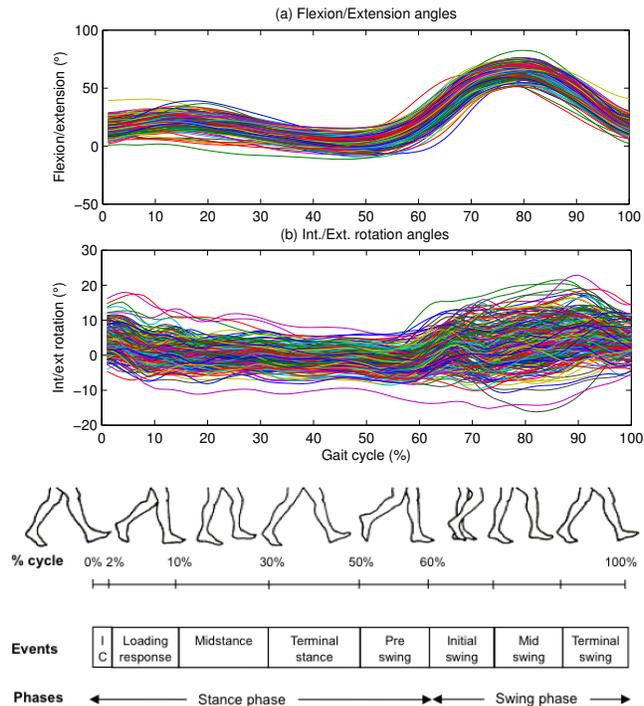
The high variability of gait data and the curse of dimensionality have forced most studies to use simple de-



**Figure 1**. Felxion/Extension (a) and Internal/External (b) angle curves over gait cycle of the database.

scriptions of the asymptomatic class, e.g., the mean of the available gait data or the simple vector of local extrema.

The characterization of kinematic data by a few representative gait patterns is basically one of data clustering because the purpose is to determine characteristic patterns in curve bundles (such as in Fig. 1). As a result, various clustering techniques have been applied including hierarchical clustering [14] , $c$-means [7, 10] and fuzzy clustering [11], using simple descriptors such as the mean of the measurement data or local determined information. How-

ever, average values and local information are, in general, ambiguous and, therefore, subjective. A preliminary study [8] showed that clustering of the global gait curve data was able to overcome this limitation. In this paper, we extend the analysis to the other two measurement planes, i.e., the flexion/extension and tibial internal/external rotation planes.

## 2 KNEE KINEMATIC CLUSTERING PROCESS

The functional diagram of the knee kinematic signals clustering process is illustrated in Fig. 2. The proposed framework consists of three steps: data dimensionality reduction, clustering of the reduced-dimensionality data and, cluster validation. We use principal component analysis to reduce the dimensionality of the original knee angle data to a few descriptive variables (Section 2.2). Doing so reveals that the sign of the principal components form clusters, which provides an effective and computationally efficient method to determine distinct patterns descriptive of the asymptomatic gait (Section 2.3). The determined clusters are verified by silhouette width and clinical interpretation (Section 2.4).

### 2.1 Kinematic data collection

Three-dimensional knee kinematics data were recorded while the participant was walking on a conventional treadmill at a self-selected comfortable speed. Treadmill kinematic data have been shown to be similar to overground gait data [13] but are much easier to obtain. Data collection was performed on each knee separately. Each participant took part in a 10 minutes treadmill walking adaptation period to ensure reproducible knee kinematics [3] prior to data acquisition. A knee marker attachment system, developed to reduce skin motion artifact [6], was installed on the participant's knee to record the 3D kinematics. The position and orientation of the markers were recorded using an electromagnetic motion tracking system (Fastrack, Polhemus, USA) at a sampling frequency of 60 Hz. A number of representative gait cycles, generally fifteen, were averaged to obtain a mean pattern per subject. This was followed by interpolation and resampling from $1\%$ to $100\%$ of the gait cycle, therefore giving a 100 measurement points for each participant (Fig.1). These data processing steps were performed with Matlab 2008b (The Mathworks, USA). A total of 111 asymptomatic volunteers participated in this gait study. For 9 of the participants, measurements were only collected on one knee, giving a total of 213 data set vectors.

### 2.2 Dimensionality reduction

Principal component analysis is an efficient technique for dimensionality reduction. It has been used in various fields of study, including biomechanics [4, 9]. Given a set of data of $p$ dimensions, PCA determines a linear subspace of dimension $k \leq p$ such that the projection of the data points on this subspace contains most of the variability of the data, i.e., most of the relevant information in the data. More precisely, let $\mathbf{X}_{(n \times p)} = \{x_{i,j} : 1 \leq i \leq n,\ 1 \leq j \leq p\}$ be the observed data presented in matrix form. In this study, each individual contributes two rows in this matrix $\mathbf{X}$, one corresponding to the left knee angle and the other to the right knee angle. Therefore, $n$ is even and $n/2$ is the number of individuals. Each column of the matrix corresponds to a gait cycle percentage. There are 100 column ($p = 100$) in our study, this corresponding to the 100 points of the angle vs. cycle percentage curve. The observation matrix $\mathbf{X}$ is generally expressed as a mean normalized to 0 and unit variance data. PCA produces linear combinations of the observations $\mathbf{X}_{(n \times p)}$ to determine the principal components (PC) scores $\mathbf{Z}_{(n \times p)}$:

$$\mathbf{Z}_{(n \times p)} = \mathbf{X}_{(n \times p)} \times \mathbf{A}_{(p \times p)} \qquad (1)$$

A powerful property of PCA is that the majority of the data variation can often be explained by just the first few principal components. As a result, the components corresponding to the smaller eigenvalues are dropped, affording a significant reduction in dimensionality without loss of relevant information. The number of principal axes to retain can be determined using the portion of explained variance criteria. This test involves plotting the eigenvalues in descending order of their magnitude against their axis numbers and determining where they level off. The break between the steep sloped portion and the levelling off portion of the curve indicates the number of meaningful principal axes.

### 2.3 Kinematic data clustering

Clustering divides a data set into groups such that the elements within a group are more similar to each other than they are to the elements of the other groups. The dimensionality reduction is followed by a discriminant model analysis, based on the PC sign, to separate the individuals gait into homogenous groups. This model provides an effective and computationally efficient method to determine distinct patterns descriptive of the asymptomatic gait. Moreover, it is consistent with the assumption of Gaussian distributed data which served the derivation of the PCA, namely to determine an orthonormal basis for the PCs [1].

### 2.4 Cluster validation: silhouette width

The purpose of cluster validation is to determine whether clustering resulted in well distinguished groups. We verified the clustering quality by silhouette width. Silhouette width is a confidence indicator of the membership of a sample to a cluster . For each cluster $C_j$, $j = 1, 2, ..., c$, the silhouette width $s(i)$ of a sample $\mathbf{x}_i$ is:

$$s(i) = \frac{b_i - a_i}{\max(a_i, b_i)}, \qquad (2)$$

where $a_i$ is the average distance between $\mathbf{x}_i$ and the elements in the cluster $C_j$, and $b_i$ is the minimum of the aver-
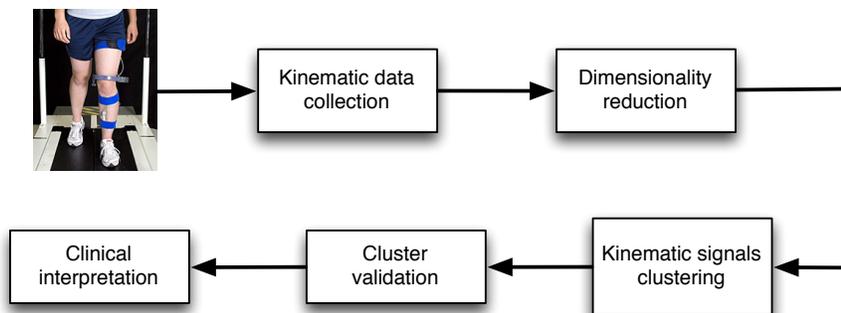
**Figure 2**. The functional diagram of the knee kinematic signals clustering process

age distances between the sample $\mathbf{x}_i$ and all of the samples clustered in $C_k$ ($k = 1, 2, ..., c$ and $k \neq j$). From this formula it follows that $-1 \leq s(i) \leq 1$. A positive value of $s(i)$ indicates that the sample has been *well-clustered*, i.e., it was assigned an appropriate cluster. If $s(i)$ is negative, the sample has been *misclassified* [2].

## 3 EXPERIMENTAL RESULTS

We implemented all aspects of kinematic dimentionality reduction, clustering and validation described in the previous sections.

Before reducing the kinematic data dimension, we verified that the PCA is appropriate to reduce the dimensionality of the original data. Using the Kaiser-Meyer-Olkin (KMO) index [12], we verified that the data sampling is adequate to conduct a PCA. The computed KMO values are $0.92$ and $0.98$ for the flexion/extension and the internal/external rotation signals respectively. Such values are considered excellents and indicate that the sampling is adequate and, therefore, a PCA is appropriate to reduce the dimensionality of the original data.

### 3.1 Dimensionality reduction and clustering

We used the portion of explained variance criteria to determine the proper number of principal components to retain. The first two PCs have been retained for each of kinematic signals, i.e., the flexion/extension and the internal/external rotation, as they explain, respectively, $70\%$ and $81\%$ of the total variance.

The dimensionality reduction is followed by a discriminant model analysis the purpose of which is to separate the individuals gait data into homogenous groups. The application of this model resulted in two clusters for each kinematic data plane. The two clusters identified in the sagittal plane are shown in Fig. 3. Each curve in a cluster corresponds to a subject flexion/extension in the original data space (Fig. 1 (a)).

The same analysis was performed in the transverse plane for the internal/external rotation angles. We obtained the two clusters represented in Fig. 4. A curve in a cluster corresponds to a flexion/extension curve subject in the original data space of Fig. 1 (b).
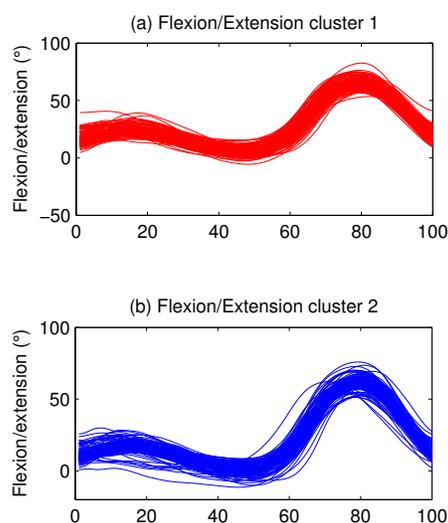


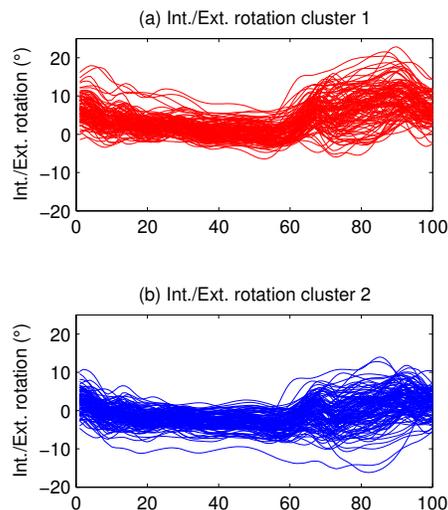**Figure 3**. The two determined flexion/extension clusters.



**Figure 4**. The two determined internal/external rotation clusters.

We note that the extracted clusters have a comparable number of samples: 51% for Cluster 1 and 49% for Cluster 2 in flexion/extension angles observations, and 50% for Cluster 1 and 50% for Cluster 2 in the internal/external angles observations. This is a desirable result because there is a priori no justification, statistical or clinical, for uneven repartition of the data points among the clusters.

### 3.2 Cluster validation

Fig. 5 shows the silhouette plot for each of the 213 flexion/extension (Fig. 5 (a)) and internal/external (Fig. 5 (b)) angles observations. The observations are grouped per cluster starting with cluster 1 at the top. The silhouette values as horizontal bars are ordered to fit in each cluster.

The silhouette width is positive for all the observations except for one sample for the flexion/extension. Three silhouette values are negative for the internal/external rotation (2 in Cluster 1 and 1 in Cluster 2) which represent only about 1% of the data set, i.e., 1% of the gait data sample have assigned to a wrong cluster in the internal/external rotation angles. These results shows a good separation quality of the clusters.
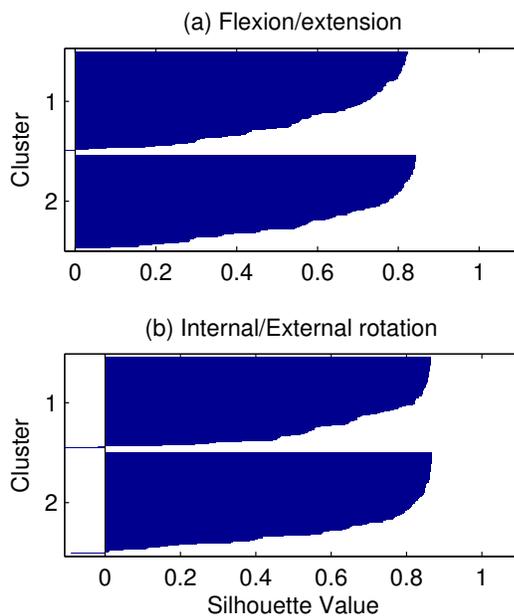


**Figure 5**. Silhouette index (a) Felxion/extension angle curves (b) Internal/external angle curves. For each data set element $\mathbf{x}_i$, $i = 1, 2, ..., 213$, a horizontal bar represent the silhouette width $s(i)$ to the cluster $C_j$, $j = 1, 2, 3, 4$. The width of a particular bar is proportional to the compactness. A negative silhouette value means that the observation is assigned to a wrong cluster.

### 3.3 Clinical interpretation

The mean of each cluster identifies a meaningful knee gait pattern in flexion/extension and tibial inter-

nal/external rotation gait as shown in Fig. 6. The analyses of the flexion/extension patterns of the Fig. 6 (a) show that the extremum amplitudes of the two clusters are observed at the same times: the maximum of the stance phase (14% of the gait cycle), minimum of the stance phase (50% of the gait cycle), and the maximum the swing phase (80% of the gait cycle). However, a shift 6° is observed at the initial contact (1% of the gait cycle) and during the stance phase (1% − 60%). The offset decrease during the swing phase.

The two average curves of internal/external rotations (Fig. 6 (b)) are much more offseted. The local extremum amplitudes of the two clusters are observed at the same percentage of the gait cycle (local maximum at 13%, 50%, 70%, 73%, 92% of the gait cycle). 7° offset occurs from the initial contact and is almost constant throughout the gait cycles. The first pattern (red curve in Fig. 6 (b)) demonstrates that individuals belonging to this cluster have a more rotated knee during the swing phase than the second pattern (blue curve in Fig. 6 (b)).
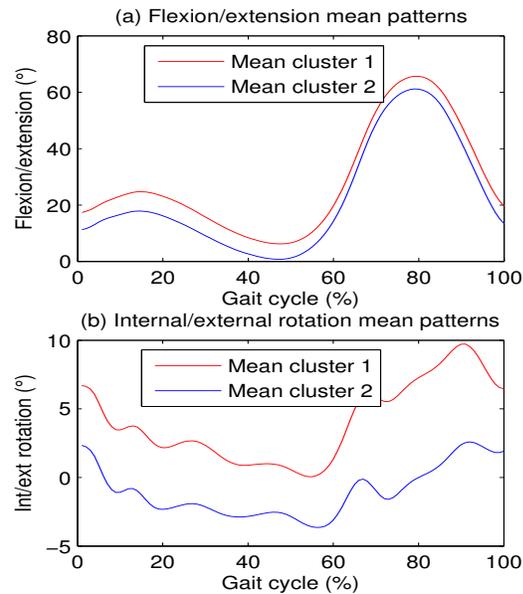


**Figure 6**. The mean patterns of the clusters.

## 4 CONCLUSION

In this study we developed a method which identified two distinct normal gait patterns by PCA of knee flexion/extension and internal/external angle data. The patterns were validated by statistical analysis via the data silhouette width, and also by clinical interpretation. For further understanding, the study can be extended to a joint analysis of the sagittal plane (tibial internal/external rotation), transverse plane (flexion/extension), and frontal plane (abduction/adduction) kinematic data.

## References

[1] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 1st ed. 2006. corr. 2nd printing edition, 2007.

[2] N. Bolshakova and F. Azuaje. Cluster validation techniques for genome expression data. *Signal Process.*, 83(4):825–833, 2003.

[3] M. Van de Putte, N. Hagemeister, N. St-Onge, G. Parent, and J.A. de Guise. Habituation to treadmill walking. *Journal of Bio-Medical Materials and Engineering*, 16(1):43–52, 2006.

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

[5] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.

[6] N. Hagemeister, G. Parent, M. Van de Putte, M. St-Onge, N. Duval, and J. A. de Guise. A reproducible method for studying three-dimensional knee kinematics. *Journal of Biomechanics*, 38(9):1926–1931, 2005.

[7] G. Kienast, D. Bachmann, G. Steinwender, E.B. Zwick, and V. Saraph. Determination of gait patterns in children with cerebral palsy using cluster analysis. *Gait & Posture*, 10:57, 1999.

[8] N. Mezghani, A. Fuentes, N. Gaudreault, A. Mitiche, N. Hagemeister, R. Aissaoui, and J.A De Guise. Identification of knee frontal plane kinematic patterns in normal gait by principal component analysis. *Journal of Mechanics in Medicine and Biology*, 13(3):17 – 24, 2013.

[9] A.M.S Muniz and J. Nadal. Application of principal component analysis in vertical ground reaction force to discriminate normal and abnormal gait. *Gait and Posture*, 29(1):31–35, 2009.

[10] J.M. O'Byrne, A. Jenkinson, and T.M. O'Brien. Quantitative analysis and classification of gait patterns in cerebral palsy using a three-dimensional motion analyzer. *Journal of Child Neurology*, 13:101–108, 1998.

[11] M.J O'Malley, M.F Abel, D.L. Damiano, and C.L Vaughan. Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters. *IEEE transactions on rehabilitation engineering*, 5(4):300–309, 1997.

[12] M.A. Pett, N.R. Lackey, and J.J. Sullivan. *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. 2003.

[13] P.O. Riley, G. Paolini, U.D. Croce, K.W. Paylo, and D.C. Kerrigan. A kinematic and kinetic comparison of overground and treadmill walking in healthy subjects. *Gait and Posture*, 26(1):17 – 24, 2007.

[14] B. Toro, C.J. Nester, and P.C. Farren. Cluster analysis for the extraction of sagittal gait patterns in children with cerebral palsy. *Gait & Posture*, 25(2):157–165, 2007.

[15] M.W. Whittle. *An Introduction to Gait Analysis*. Butterworth-Heinemann, 2007.