

Identification des patrons de la cinématique du genou par une analyse en composantes principales

Identification of knee kinematic patterns in normal gait by principal component analysis

N. Mezghani^{1,2}, N. Gaudreault³, A. Fuentes¹, A. Mitiche⁴,
N. Hagemeister¹, R. Aissaoui¹, et J.A. de Guise¹

¹ Laboratoire de recherche en imagerie et orthopédie
Centre de recherche du CHUM, École de technologie supérieure
Montréal, Québec

neila.mezghani@etsmtl.ca

² UER Science et Technologie, TÉLUQ, Montréal, Québec

³ École de réadaptation, Université de Sherbrooke,
Sherbrooke, Québec, Canada.

⁴ Institut National de la Recherche Scientifique, INRS-EMT
Montréal, Québec, Canada.

Résumé Le but de cette étude est d'identifier les différents patrons de la cinématique du genou d'une grande population d'individus asymptomatiques. Les données cinématiques considérées sont enregistrées dans le plan frontal. Elles correspondent au mouvement d'abduction/adduction durant le cycle de marche. La méthode proposée utilise une analyse en composantes principales (ACP) qui réduit la dimensionnalité des données en les projetant dans un sous-espace de composantes principales significatives (CP). La réduction de la dimensionnalité est suivie par un regroupement selon un modèle discriminant qui permet de séparer les données cinématiques en groupes homogènes d'individus. Quatre types de patrons descriptifs ont été identifiés et validés par la silhouette des classes ainsi que par des tests statistiques d'hypothèses.

Mots clés Regroupement, Analyse en composantes principales, Cinématique du genou, Patrons de la cinématique, Analyse de la marche ;

Abstract The purpose of this study is to identify meaningful gait patterns in knee frontal plane kinematics from a large population of asymptomatic individuals. The proposed method uses principal component analysis (PCA). It first reduces the data dimensionality, without loss of relevant information, by projecting the original kinematic data onto a subspace of significant principal component (PCs). This is followed by a discriminant model to separate the individuals gait into homogenous groups. Four descriptive gait patterns were identified and validated by clustering silhouette width and statistical hypothesis testing, as well as a clinical interpretation.

Key words Clustering, Principal component analysis, Knee kinematic data, Kinematic patterns, Gait analysis.

1 Introduction

L'analyse de la marche à partir de données cinématiques du genou (angles de flexion/extension, d'abduction/adduction et de rotation interne/externe) offre des informations sur la fonction de locomotion humaine. Une caractérisation par des patrons représentatifs des données d'une population asymptotique, aussi appelée population normale, est importante pour mieux comprendre cette fonction [17]. Cependant, une telle caractérisation est difficile pour deux raisons. Premièrement, les données cinématiques de chaque sujet sont fournies sous la forme d'un vecteur de mesures de grande dimension rendant l'analyse vulnérable à la malédiction de la dimensionnalité [5]. Deuxièmement, ces données présentent une grande variabilité. Plusieurs études ont attribué la variabilité de ces données à la différence de méthodologie d'acquisition des données (systèmes d'acquisition et de traitement des données, vitesse de marche, etc) [4]. D'autres études ont émis l'hypothèse de la présence de plusieurs patrons de données cinématiques [18].

L'identification des patrons représentatifs à partir d'un grand ensemble de données non étiquetées est un problème de regroupement (*clustering* en anglais). Plusieurs études ont appréhendé ce problème pour l'identification de patrons caractéristiques de données cinématiques de la locomotion humaine par diverses méthodes, par exemple par le regroupement hiérarchique [14,15,16], par l'algorithme des *c*-moyennes [8,10,11,19] et par la classification floue [12]. Cependant, ces études ont généralement porté sur l'analyse des données d'une population pathologique plutôt que dans une population asymptotique. Ceci est dû essentiellement au manque de données provenant de populations asymptotiques [3]. De plus, ces études ont souvent utilisé des descriptions simples basées sur des informations locales (maximums et minimums locaux) ce qui limite la portée de l'analyse.

Pour une interprétation significative des données cinématiques du genou durant la marche, nous avons utilisé des données descriptives globales plutôt que locales, à savoir sur toute la courbe des données durant un cycle entier de marche. Nous nous sommes intéressées aux données du plan frontal correspondant à l'angle d'abduction/adduction (abd/add) qui est généralement considéré comme le plan le plus représentatif. La méthode que nous proposons s'effectue en deux étapes: la dimension des données est d'abord réduite en projetant dans un sous-espace de composantes significatives par une analyse en composantes principales (ACP). Un modèle discriminant est ensuite utilisé pour séparer les individus en groupes homogènes.

Cet article est organisé comme suit: la section 2 explique l'acquisition des données et leur prétraitement. La Section 3 décrit la méthode proposée pour l'identification des patrons cinématiques ainsi que leurs critères de validation. La section 4 donne les résultats et la section 5 contient une conclusion.

2 Acquisition des données

Les données cinématiques du genou dans l'espace 3D ont été enregistrés lors de la marche à une vitesse confortable sur un tapis roulant. La collecte des données a été effectuée sur chaque genou séparément en utilisant des marqueurs de mouvement fixés au genou à l'aide du système d'attache KneeKG [6,9]. La position et l'orientation des marqueurs sont enregistrées par un système électromagnétique de suivi de mouvement (Fastrack, Polhemus, USA) dont la fréquence d'échantillonnage est de 60 Hz.

Cent onze (111) individus asymptotiques ont participé à cette étude. Les caractéristiques démographiques descriptives de ces participants sont présentées dans le tableau 1. Pour 9 des participants, des mesures ont été collectées que sur un seul genou, ce qui donne un total de 213 courbes d'abd/add (Figure 1). L'acquisition des données est suivie d'un prétraitement qui consiste en (1) un filtrage pour supprimer le bruit dû aux artefacts d'acquisition de données (2) un ré-échantillonnage entre 0 et 100% (0% correspondant au début du cycle de marche et 100% à la fin) pour normaliser la dimension des données à 100 points pour chaque individu (3) Suppression de onze courbes de données aberrantes (*outliers*) en utilisant la distance de Mahalanobis, ce qui donne un total de 202 courbes.

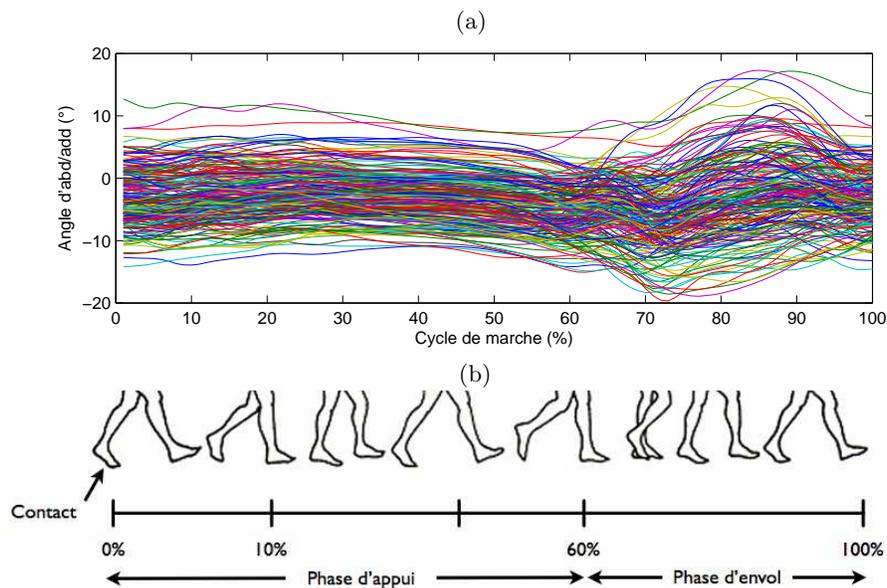


Figure 1. (a) Angle d'abduction/adduction de 213 individus asymptotiques durant le cycle de marche: les courbes ont été normalisées entre 1% et 100% (100 points) du cycle de marche. (b) Les phases du cycle de marche.

3 Identification des patrons

3.1 Analyse en composante principale

Soit un ensemble de données de dimensions p , l'analyse en composantes principales (ACP) détermine un sous-espace vectoriel de dimension $k \leq p$ tel que la projection des données sur ce sous-espace contient la plupart des l'information pertinente des données originales.

Plus précisément, soit $\mathbf{X}_{(n \times p)} = \{x_{i,j} : 1 \leq i \leq n, 1 \leq j \leq p\}$ l'ensemble des données observées présentés sous la forme d'une matrice. Dans cette étude, chaque individu est décrit par deux lignes dans la matrice \mathbf{X} , une correspondant à l'angle d'abd/add du genou gauche et l'autre du genou droit. Chaque colonne de la matrice correspond à un pourcentage du cycle de marche, ce qui qui donne 100 colonnes ($p = 100$). La matrice d'observation $\mathbf{X}_{(n \times p)}$ est généralement réduite à la variance unitaire et centralisée à une moyenne nulle. L'ACP donne un nouvel ensemble de variables $\mathbf{Z}_{(n \times p)}$ ou composantes principales (CP) qui sont des combinaisons linéaires des variables originales.

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

La matrice \mathbf{Z} , qui contient les CP ainsi générées, a la même dimension que la matrice des données originales ($n \times p$). Cependant, afin de réduire la dimensionnalité des données un nombre de composantes principales $k \leq p$ est retenu en tenant compte de la variation inhérente des données par rapport au maximum d'information possible. Le nombre k est déterminé selon le critère de la variance expliquée [7] que nous avons fixé à 90%. Les CP retenues sont utilisées par la suite pour le regroupement.

3.2 Regroupement

Le regroupement (clustering) divise l'ensemble de données en groupes de telle sorte que les éléments à l'intérieur d'un même groupe sont plus proches qu'ils ne le sont pour les éléments des autres groupes. Plus la similarité (ou homogénéité) au sein d'un groupe est grande, meilleure est la qualité du regroupement.

Le regroupement des données a été effectué selon un modèle discriminant basé sur le signe des CP. Ce modèle fournit une méthode simple et efficace pour identifier les patrons de la cinématique genou. De plus, il est compatible avec l'hypothèse de normalité des données ayant servi de base au principe de l'ACP [2]. La validité et l'efficacité d'une telle division typologique des données sont évaluées comme décrit à la prochaine section.

3.3 Validation des patrons

Les groupes déterminés ont été validés par l'indice de silhouette ainsi que par des tests statistiques d'hypothèses.

L'indice de silhouette est un indicateur efficace pour valider une méthode de regroupement [13]. Pour tout individu i de l'ensemble des données, l'indice de silhouette est défini par la formule suivante:

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

où a_i est la dissimilarité moyenne entre l'individu i et tous les autres individus du groupe C_j auquel il appartient, et b_i est le minimum des dissimilarités moyennes entre l'individu i et tous les autres individus des groupes C_k ($k = 1, 2, \dots, c$ avec $k \neq j$).

Une analyse de variance a été également effectuée afin de vérifier l'homogénéité des groupes déterminés. Un test post hoc de Tukey a été utilisé pour examiner les différences entre les paires de groupes. L'analyse statistique a été effectuée en utilisant le logiciel d'analyse statistique SPSS 18.0 (Statistical Package for Social Sciences) [1]. Un seuil de $P = 0.05$ a été fixé pour démontrer que les différences entre les groupes sont statistiquement significatives.

4 Résultats

Nous avons utilisé le pourcentage de la variance expliquée comme critère pour déterminer le nombre de composantes principales à retenir. Les deux premières composantes principales (CP) ont été retenues, car elles représentent 94% de la variance totale. La première composante, CP1, représente 79% et la deuxième composante, CP2, représente 15%.

Le regroupement des CP a permis d'obtenir les 4 groupes d'individus illustrés dans la Figure 2. À partir des courbes de chaque groupe, une courbe moyenne est produite pour obtenir les patrons descriptifs de la cinématique du genou (Figure 3). Ce résultat a une importance clinique parce qu'il permet de conclure qu'il existe quatre patrons d'abd/add du genou lors de la marche: le premier patron est proche de la position neutre (autour de 0°) au cours de la phase d'appui et en adduction (angle $> 0^\circ$) pendant la phase d'envol (Patron 1). Le deuxième patron est en abduction (angle $< 0^\circ$) pendant la phase d'appui et tend à l'adduction pendant la phase d'oscillation (Patron 2). Le troisième modèle est proche de la position neutre au cours de la phase d'appui et en abduction pendant la phase d'oscillation (Patron 3) et le quatrième patron est en abduction (angle $< 0^\circ$) pendant les phases d'appui et d'oscillation (Patron 4).

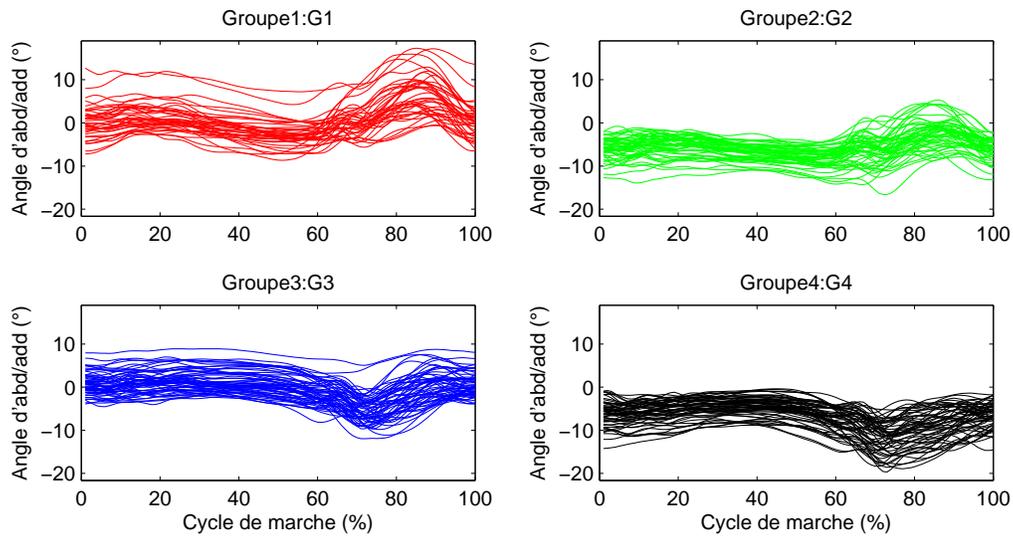


Figure 2. Les courbes correspondant à chaque regroupement: G1 à G4.

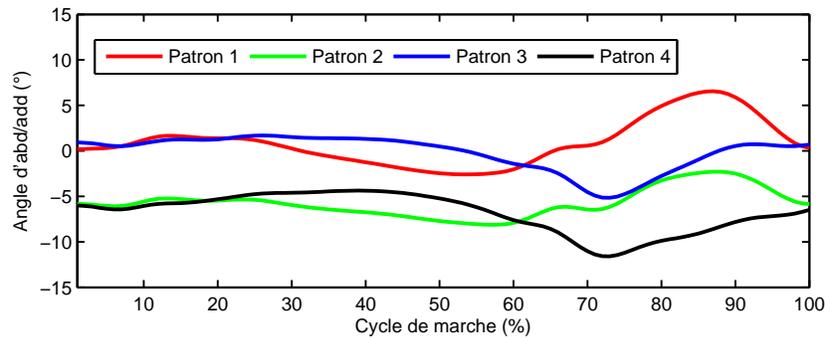


Figure 3. Les patrons de la cinématique du genou représentant une population asymptotique

La représentation graphique de l'indice de silhouette (Fig. 4) montre une bonne séparation en 4 groupes des 213 courbes d'abd/add. Ces groupes ont des nombres d'éléments similaires (21 % pour le groupe 1, 28 % pour le groupe 2, 28 % pour le groupe 1, et 27 % pour le groupe 4). Ce résultat est satisfaisant car il n'y a *a priori* aucune justification statistique ou clinique sur la répartition inégale des échantillons de la base de données entre les différents groupes. L'indice de silhouette (Fig. 4) est positif pour la majorité des échantillons, sauf pour 6 observations (1 appartenant à G1 et 5 à G2). Ceci démontre l'homogénéité des données au sein des groupes puisque seulement 3% (6/202) des échantillons ont été affectés à un groupe qui n'est pas le leur.

Identification des patrons de la cinématique du genou

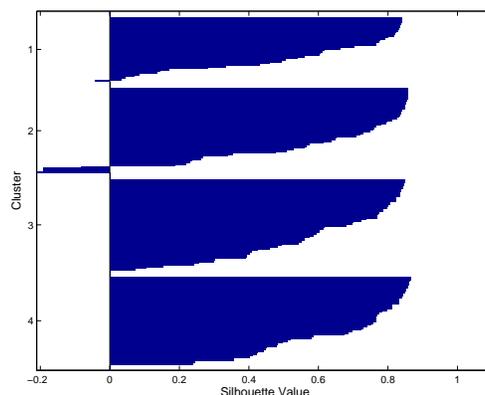


Figure 4. Indice de silhouette des données regroupées.

Les résultats des tests statistiques du Tableau 1 montre qu'il n'existe pas de différences significatives pour la vitesse de marche entre les quatre groupes. Ils montrent également l'homogénéité des caractéristiques anthropométriques pour les 3 groupes G1, G2 et G3. Ces résultats mettent en évidence que les groupes déterminés par le modèle discriminant sont indépendants de l'âge, du poids et de la hauteur des individus.

Table 1. Les caractéristiques anthropométriques de l'ensemble des sujets de la base de données et des différents groupes déterminés. (* $p < 0.05$ entre (C3 et C4). ** $p < 0.01$ entre (C1 et C4), (C2 et C4) et (C3 et C4)).

	L'ensemble des données	Groupe 1: G1	Groupe 2: G2	Groupe 3: G3	Groupe 4: G4
Number of samples	202	40	52	56	54
Age (year)	35.9±14.2	39.1 ±17.8	34.7 ±13.0	39.2 ±15.0 *	31.6 ±10.1 *
Poids (kg)	71.7 ±13.5	74.8 ±10.4 **	74.1 ±16.1 **	73.6 ±13.8**	65.1 ±10.5 **
Hauteur(cm)	169.7 ±8.5	171.2±8.4	168.5±8.9	170.2±8.9	169.3±7.9
Vitesse de marche(m/s)	1.1 ±0.1	0.99 ±0.1	1.1±0.2	1.0±0.2	1.1±0.1
Pourcentage de male/femelle	48/52 %	56/44 %	53/47%	42/58%	43%

5 Conclusion

Dans cette étude, nous avons identifié quatre patrons de la cinématique du genou d'une population asymptotique. L'identification de ces patrons est importante parce que ces derniers peuvent servir de référence pour l'analyse des données cinématiques provenant d'une population pathologique. L'utilisation de toute la

courbe a permis de garder toute l'information contenue dans la cinématique du genou lors de la marche. Cette étude peut être généralisée à l'analyse des données cinématiques du plan sagittal et transversal afin de déterminer des patrons des angles de flexion/extension et de rotation interne/externe.

Remerciements

Les auteurs tiennent à remercier le programme MITACS et le Fonds de recherche du Québec - Nature et technologies (FRQNT), le conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) et la Chaire de recherche du Canada en imagerie 3D et ingénierie biomédicale. Ils tiennent à remercier, également, Gerald Parent et Mathieu Charbonneau pour leur soutien technique.

References

1. D. Altman. *Practical statistics for medical research*. New York, NY: Chapman & Hall, 1991.
2. C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 1st ed. 2006. corr. 2nd printing edition, 2007.
3. E.Y. Chao, R.K. Laughman, E. Schneider, and R.N. Stauffer. Normative data of knee joint motion and ground reaction forces in adult level walking. *Journal of Biomechanics*, 16(3):219–233, 1983.
4. T. Chau, S. Young, and S. Redekop. Managing variability in the summary and comparison of gait data. *Journal of NeuroEngineering and Rehabilitation*, 22(2):22–42, 2005.
5. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.
6. N. Hagemeister, L'H. Yahia, N. Duval, and J.A. de Guise. In vivo reproducibility of a new non-invasive diagnostic tool for three-dimensional knee evaluation. *The Knee*, 6(3):175 – 181, 1999.
7. J. Edward Jackson. *A User's Guide to Principal Components*. John Wiley & Sons, New York, 1991.
8. 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.
9. S. Lustig, R. Magnussen, L. Cheze, and P. Neyret. The kneekg system: a review of the literature. *Knee Surgery, Sports Traumatology, Arthroscopy*, pages 1–6, 2011. 10.1007/s00167-011-1867-4.
10. S. Mulroy, J. Gronley, W. Weiss, C. Newsam, and J. Perry. Use of cluster analysis for gait pattern classification of patients in the early and late recovery phases following stroke. *Gait & Posture*, (18):114–125, 2003.
11. 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.
12. 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.
13. J.P. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20(1):53–65, November 1987.
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. V.G. Vardaxis, P. Allard, R. Lachance, and M. Duhaime. Classification of able-bodied gait using 3-d muscle powers. *Human Movement Science*, 17(1):121–136, 1998.
16. E. Watelain, F. Barbier, P. Allard, A. Thévenon, and J.C. Angué. Gait pattern classification of elderly healthy men on the basis of biomechanical data. *Archives of physical medicine and rehabilitation*, 5(81):579–586, 2000.
17. M.W. Whittle. *An Introduction to Gait Analysis*. Butterworth-Heinemann, 2007.
18. D.A. Winter. Kinematic and kinetic patterns in human gait: variability and compensating effects. *Human Movement Science*, 3(1-2):51–76, 1984.
19. G. Xu, Y. Zhang, and R. Begg. Mining gait pattern for clinical locomotion diagnosis based on clustering techniques. *Advanced Data Mining and Applications*, pages 296–307, 2006.