Title: Defining gait patterns using Parallel Factor Analysis (Parafac2): a new analysis of previously published data.

Article type: Original full length article

Bernard X.W. Liew, PhD [a]

Susan Morris, PhD [b]

Kevin Netto, PhD [b]

[a] Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, UK.

[b] School of Physiotherapy and Exercise Sciences, Curtin University, GPO Box U1987, Perth, WA 6845, Australia.

Corresponding author: Mr Bernard Liew, School of Sport, Exercise and Rehabilitation Sciences, Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), College of Life and Environmental Sciences, University of Birmingham, Birmingham, UK; E-mail: LiewB@adf.bham.ac.uk; liew_xwb@hotmail.com. Tel: +44 121 415 8526

Abstract word count: 250

Word count: 2500

Figures and Table: 3 figures and 2 table

Supplementary material: 6

Abstract

Three-dimensional gait analysis (3D–GA) is commonly used to answer clinical questions of the form "which joints and what variables are most affected during when". When studying high-dimensional datasets, traditional dimension reduction methods (e.g. Principal Components Analysis) require "data flattening", which may make ensuing the solutions difficult to interpret. The aim of the present study is to present a case study of how multi-dimensional dimension reduction technique, Parallel Factor 2 (Parafac2), provides a clinically interpretable set of solutions to typical biomechanical datasets where different variables are collected during walking and running. Three-dimensional kinematic and kinetic data used for the present analyses came from two publicly available datasets on walking (n = 33) and running (n= 28). For each dataset, a four-dimensional array was constructed as follows: Mode A was time normalized cycle points; mode B was the number of participants multiplied by the number of speed conditions tested; mode C was the number of joint degrees of freedom, and mode D was variable (angle, velocity, moment, power). Five factors for walking and four factors for running were extracted which explained 79.23% and 84.64% of their dataset's variance. The factor which explains the greatest variance was swing-phase sagittal plane knee kinematics (walking), and kinematics and kinetics (running). Qualitatively, all extracted factors increased in magnitude with greater speed in both walking and running. This study is a proof of concept that Parafac2 is useful for performing dimension reduction and producing clinically interpretable solutions to guide clinical decision making.

Key Words: Walking, Running, Kinematics, Kinetics, Multivariate statistics

1. Introduction

Three-dimensional gait analysis (3D–GA) is commonly used in clinical biomechanics for guiding clinical decision making (Steele et al., 2015). This form of analysis produces significant amounts of data which can make clinical utilization challenging. Dimension reduction, such as principal components analysis (PCA) (Schwartz and Rozumalski, 2008), is commonly applied to biomechanical data to distill complex multivariate biomechanical variables into simpler univariate variables. For dimension reduction to generate clinically interpretable solutions, it needs to answer clinical questions in the form of "which joints and what biomechanical variables are most affected during which phase of a locomotion cycle".

Existing dimension reduction methods (e.g. PCA) may not provide interpretable clinical answers, as they require "data flattening". Biomechanics data can have up to four dimensions (4D), such as when studying many participants (dimension 1), many time points within a movement cycle (dimension 2), many joint degrees of freedom (DOF) (dimension 3), and many types of biomechanical variables (dimension 4). To process a 4D dataset using PCA, the input data must be flattened into 2D (Schwartz and Rozumalski, 2008), which confounds the variability of one dimension with another. For example, flattening a 9 x 51 (angle by time) matrix into a 459 x 1 vector combines inter-angle with inter-time variability, which precludes knowing which joint angle is affected at what locomotion phase.

To retain the underlying dimensional structure of biomechanical data, dimension reduction can be performed using techniques such as Parallel Factor Analysis (Parafac) (Harshman, 1970; Harshman and Lundy, 1994) and its extension Parafac2 (Harshman, 1972). Parafac is a multi-dimensional extension of PCA which makes it possible to extract clinically interpretable factors that uniquely differentiate individual and/or group motions across a movement cycle and between different joint DOF (Helwig et al., 2012). Parafac2 extends the functionality of Parafac to handle biomechanical variables of different units of measure (Helwig et al., 2013; Liew et al., 2018).

The primary aim of this study is to present a case study demonstrating the potential for Parafac2 to provide a clinically interpretable set of solutions to typical biomechanical datasets for walking and running. We hypothesized that the factor which contributes to the greatest variance accounted for (VAF) in walking will have a significant contribution from the knee joint during the swing phase (Boccia et al., 2018); while in running, the VAF will be from the knee joint during the stance phase (Saito et al., 2018).

2. Methods

2.1. Experimental set up

Data for the current study came from two publicly available datasets on walking (n = 42) and running (n = 28) in healthy adults (Fukuchi et al., 2018; Fukuchi et al., 2017). The detailed methods of the studies can be found in the primary open source publications (Fukuchi et al., 2018; Fukuchi et al., 2017).

<u>Walking data:</u> Participants performed unshod walking on a dual-belt, force-instrumented treadmill (300 Hz, FIT; Bertec, Columbus, OH, USA), and motion was captured with 12 opto-electronic cameras (150Hz, Raptor-4; Motion Analysis Corporation, Santa Rosa, CA, USA) (Fukuchi et al., 2018). Walking occurred over eight controlled speeds: 40%, 55%, 70%, 85%, 100%, 115%, 130%, 145% of each participant's self-determined dimensionless speed (Froude number). The eight dimensionless speed corresponded to a mean (standard deviation [sd]) speed of 0.5 (0.1) m/s, 0.7 (0.1) m/s, 0.9 (0.1) m/s, 1.1 (0.1) m/s, 1.2 (0.2) m/s, 1.4 (0.2) m/s, 1.6 (0.2) m/s, 1.8 (0.2) m/s, respectively. Marker trajectories and ground reaction force (GRF) were collected for 30s and the data were low passed filtered at a matched frequency of 6Hz (4^{th} Order, zero-lag, Butterworth) (Kristianslund et al., 2012).

<u>Running data:</u> Running assessment was performed using the same opto-electronic cameras and force-instrumented treadmill as in walking (Fukuchi et al., 2017). Participants

performed shod running across three fixed speeds of 2.5 m/s, 3.5 m/s, and 4.5 m/s (Fukuchi et al., 2017). Marker trajectories and GRF were collected for 30s and the data were low passed filtered at a matched frequency of 12Hz (4^{th} Order, zero-lag, Butterworth) (Kristianslund et al., 2012).

2.2. Biomechanical processing

Biomechanical modelling was performed in Visual 3D software version 6.00.33 (C-motion Inc., Germantown, MD, USA). A force plate threshold of 50N was used to determine gait events of initial contact and toe-off. A seven segment lower limb model was created for each study (Fukuchi et al., 2018; Fukuchi et al., 2017). The following convention was used as the anatomical coordinate system: flexion-extension occurred about the Z-axis, with positive pointing laterally to the right; abduction-adduction occurred about the X-axis with positive pointing anteriorly; and internal-external rotation occurred about the Y-axis with positive pointing superiorly.

Three-dimensional 3D joint angle, velocity, internal moment, and power, of the right ankle, knee, and hip joints. Joint angle was calculated using a Cardan flexion-abduction-rotation sequence (Cole et al., 1993). Joint velocity and moment were expressed in the proximal segment's reference frame (Schache and Baker, 2007). Joint power was calculated by the product of joint moment and velocity. All biomechanical variables were time normalized to 101 data points, between two consecutive initial contacts of the right limb. Both joint moment and power were normalized to body mass.

2.3. Parallel factor analysis

The biomechanical waveform variables for each participant at each speed were averaged across all strides for walking and running, such that one participant had one stride for each locomotion speed. Parafac2 analysis was performed in R software (v 3.2.5) (R Core

Team (2017)), using the *"multiway"* package (Helwig, 2018). The input data was organized into a four dimensions: 1) mode A being time (101 cycle points), 2) mode B being the vectorization of participants by speed (number of participants multiplied by number of speed conditions), 3) mode C being the nine joint DOFs – three cardinal axes of three joints, and 4) mode D being the four biomechanical variables (angle, velocity, moment, power). The organization of the data for mode B enables identification of factors common to all participants and all locomotion speeds.

Each variable within mode D was scaled to have a root mean square of one:

$$\underline{X}_l^* = s_l^{-1/2} X_l, where\ s_l = (IJK)^{-1} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} x_{ijk(l)}^2 \dots (1)$$

where *l* represents each of the four variables within mode D. This allowed each biomechanical variables to contribute equally to the Parafac2 solutions (Helwig et al., 2013).

Mode B weights reflect the instance of each participant at each speed for the extracted factors. Mode A (time) was analysed simultaneously with mode D (variables), given that a Parafac2 model requires the nesting of mode A within mode D to achieve a least squared solution (Helwig et al., 2013). A higher absolute weighting indicates a greater contribution of the mode to the factor (Helwig et al., 2013; Helwig et al., 2012). Understanding the "mechanical" meaning of the extracted factors were determined by comparing modes C (DOF), and A-D (time by variables) weights, to the original biomechanical variables. For example, if knee sagittal plane in mode C and joint velocity at 85% stride in modes A-D had the highest weights in factor one, then factor one was interpreted to reflect knee flexion-extension swing phase kinematics.

An Alternating Least Squares (ALS) algorithm was used to find an optimal solution, using 500 random starts with 500 maximum iterations of the ALS algorithm for each start.

An orthogonal constraint was applied to Mode C to improve the interpretability of the Parafac2 solutions (Helwig et al., 2013). Parafac2 solutions were scaled such that mode B weights absorbed the data's scale so that the mode B weights can be used to understand differences in locomotive patterns at different speed. Mode A weights absorbed the data's sign because the time weights have a meaningful bipolar interpretation (e.g. positive for joint flexion). For each gait type, the number of factors extracted was determined when an additional factor did not increase the variance accounted for (%VAF) by > 3% (van den Hoorn et al., 2015). We plotted the mean with standard deviation of the cohort's mode B weights at each locomotion speed, to provide a description of how each factor varied its weightage with speed.

3. Results

Nine out of the 42 participants from the walking dataset were excluded from the present study. These participants had many simultaneous bilateral foot contacts on the same force plate, resulting in an absence of consecutive good foot contact strides which lasted > 50% of the walking duration. The 50% threshold was determined by the authors to minimize manual identification of foot contact events, to increase processing replicability. Summary statistics of the demographic characteristics, as well as the original biomechanical waveforms can be found in the supplementary material.

For walking, five factors were extracted which accounted for 79.23% of the VAF (Table 1, Figure 1). All factors increase their weightage, but factors 1, 2, and 4 had the greatest increase over all walking speeds (Figure 3a). For running, four factors were extracted which explained 84.64% of the VAF of the walking dataset (Table 1, Figure 2). All factors increase their weightage, but factors 1, 2, and 4 had the greatest increase over the running speeds investigated (Figure 3b). The mechanical interpretations of all extracted factors in walking and running can be found in Table 2.

4. Discussion

The aim of the present study was to illustrate how Parafac2 provides a clinically interpretable set of solutions to typical multidimensional biomechanics datasets. We found that the factor with the largest VAF in walking and running weighted highest for knee flexion-extension mechanics during the swing phase, which partially supported our hypothesis.

Swing-phase knee flexion-extension mechanics reflected the largest variance of steady-speed walking and running, suggesting that the motor control system does not exert "strong" independent control (resulting in low variability) of the knee during swing. This is not surprising given that the energetic cost of leg swing is low, compared to the stance phase (Arellano and Kram, 2014; Gottschall and Kram, 2005), and tissue loads are low in the swing phase. The large variability of swing-phase knee flexion-extension mechanics may allow inter-joint covariation to ensure adequate toe-clearance height (Srivastava et al., 2016). Ensuring adequate toe-clearance is essential to enable forward progression, whilst avoiding ground obstacles to prevent falls. Previous studies have also reported the importance of knee flexion angular velocity (Chou and Draganich, 1998), and knee flexion-extension range of motion (Benson et al., 2018) during the swing-phase of walking in determining the risk of falling potential.

If healthy walking behaviour is typified by a high variance explained by swing-phase knee biomechanics, which becomes more important at greater walking speeds, than impaired walking should be affected on both these features. Children with cerebral palsy have a reduced variance explained by the factor with a high contribution from swing-phase knee flexor muscle activity, compared to healthy children (Steele et al., 2015). The experience of inadequate swing-phase knee flexion during walking in children with cerebral palsy has been termed as “stiff knee gait” (Bohm et al., 2014). Impaired swing-phase knee flexion may be magnified with greater walking speed due to muscle spasticity (Van Campenhout et al., 2014). Adults with stroke have also displayed impaired capacity to fractionate the timing of their muscle synergy activations, particularly during the swing phase of walking (Clark et al., 2010).

In clinical biomechanics, dimension reduction has been used to develop normative databases of healthy and impaired walking behaviour in children (Schwartz and Rozumalski, 2008). The present paper provides a foundation upon which future studies can develop clinically interpretable normative databases of healthy and impaired walking, as well as running, behaviours in adults. The principle limitation of the present study was that walking and running were performed on a treadmill. Whether the present findings will be replicated during overground locomotion is unclear. However, a previous study reported similar muscle synergies extracted during running in both overground and treadmill conditions (Oliveira et al., 2016).

5. Conclusion

Factors with high weightings from the knee flexor-extensor mechanics during the swing phase accounted for the largest variance in normal walking and running, and its importance increased at faster locomotion speeds. This study is a proof of concept that Parafac2 is useful for performing dimension reduction in high-dimensional biomechanics datasets, to produce clinically interpretation solutions.

Figure 1.

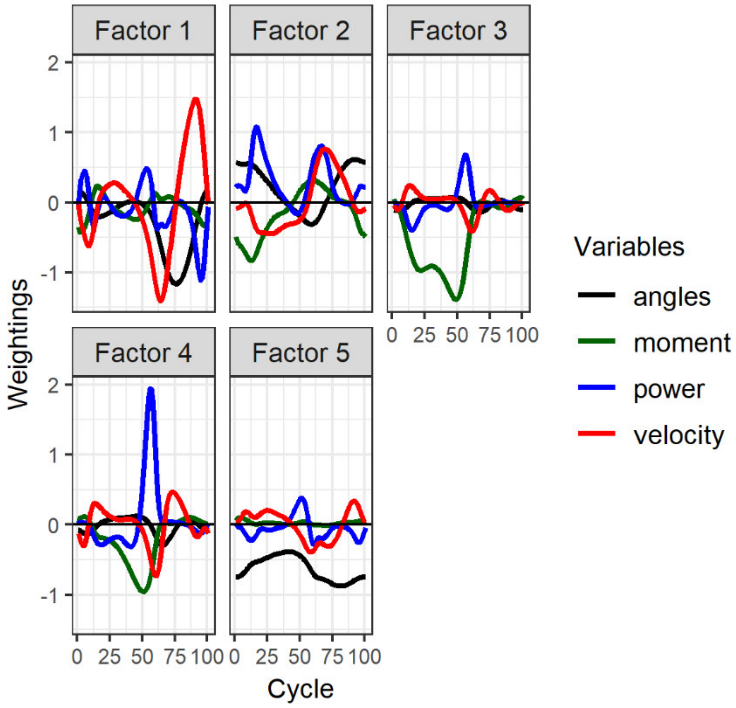

Figure 2.

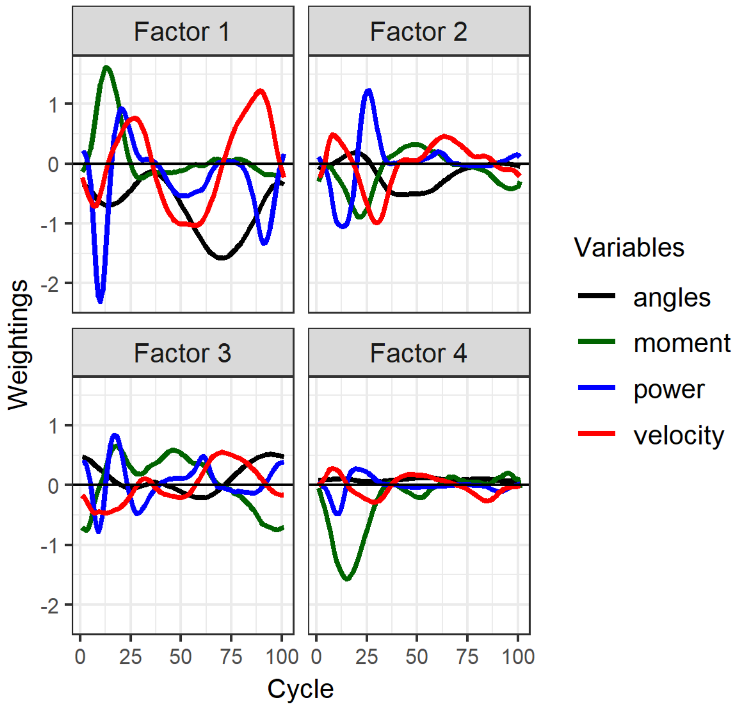

Figure 3

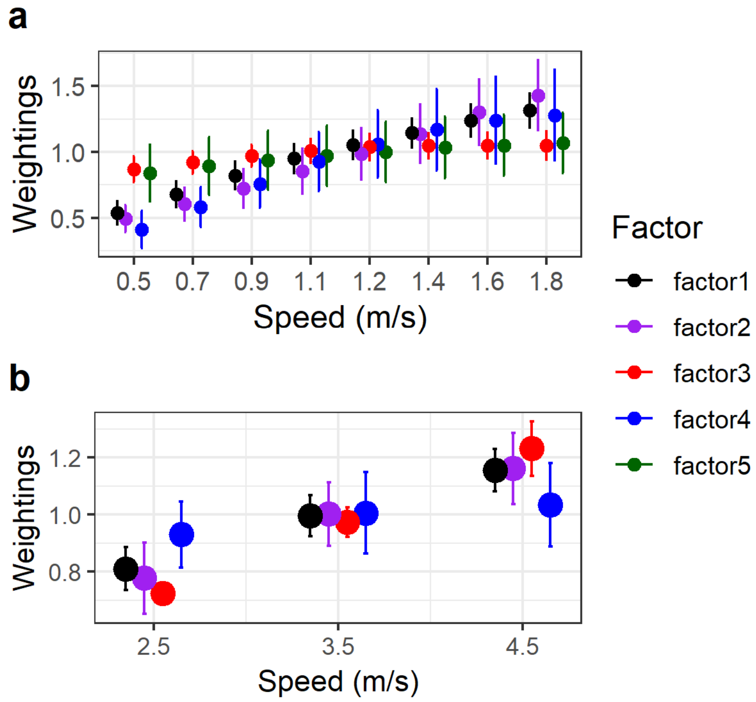
a
Weightings
1.5
1.0
0.5
0.5
0.7
0.9
1.1
1.2
1.4
1.6
1.8
Speed (m/s)
b
Weightings
1.2
1.0
0.8
2.5
3.5
4.5
Speed (m/s)
Factor
factor1
factor2
factor3
factor4
factor5

Table 1. Weightings of each degree of freedom (DOF) (mode C) for each extracted factor during walking and running

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
DOF					
Walking					
Ankle INV/EVE	0.58	-0.38	0.15	0.08	**-1.70**
Ankle ABD/ADD	0.16	-0.46	1.46	-0.65	-1.17
Ankle PF/DF	0.00	0.05	0.89	**2.85**	-0.04
Knee ABD/ADD	-0.03	0.28	1.18	-0.37	0.12
Knee IR/ER	-0.40	0.13	0.64	-0.18	**1.89**
Knee FLEX/EXT	**2.91**	0.03	0.01	-0.01	0.71
Hip ABD/ADD	0.00	0.25	**2.03**	-0.50	0.20
Hip IR/ER	-0.12	0.28	0.38	-0.09	0.50
Hip FLEX/EXT	0.10	**2.90**	-0.11	-0.04	-0.58
% VAF	35.07	21.60	13.12	5.63	3.83
Run					
Ankle INV/EVE	0.16	-0.19	0.12	-0.12	-
Ankle ABD/ADD	0.00	-0.44	0.23	0.47	-
Ankle PF/DF	0.15	**2.70**	-0.91	0.63	-
Knee ABD/ADD	0.01	0.07	0.07	1.09	-
Knee IR/ER	0.16	0.21	-0.36	0.04	-
Knee FLEX/EXT	**2.74**	0.32	1.10	-0.37	-
Hip ABD/ADD	0.27	-0.42	0.50	**2.60**	-
Hip IR/ER	-0.01	0.43	0.05	-0.52	-
Hip FLEX/EXT	-1.17	0.98	**2.55**	-0.18	-
% VAF	52.62	16.21	11.89	3.93	-

Abbreviations:
INV = inversion; EVE = eversion; ABD = abduction; ADD = adduction; PF = plantarflexion; DF = dorsiflexion; FLEX = flexion; EXT = extension; IR = internal rotation; ER = external rotation; % VAF = percentage variance accounted for

Table 2. Mechanical interpretation of extracted factors in walking and running

Factors	Mode C (DOF) with high weights	Mode A-D (time-variables) with high weights	Interpretation
Walk			
1	Knee FLEX/EXT	Velocity, angle, power after 65% stride	Knee flexion mechanics in swing
2	Hip FLEX/EXT	Power before 25% stride	Hip extension power single support
3	Hip ABD/ADD	Moment before 65% stride	Hip abduction moment stance
4	Ankle PF/DF	Power before 65 % stride	Ankle plantarflexion power at push-off
5	Knee IR/ER	Angle after 65% stride	Knee rotation kinematics in swing
Run			
1	Knee FLEX/EXT	Moment and power before 25%; Angle, velocity, power after 65% stride	Knee extension kinetics early stance; Knee flexion mechanics in swing
2	Ankle PF/DF	Power before and after 25% stride	Hip extension power single support
3	Hip FLEX/EXT	No dominant biomechanical variables	Hip sagittal plane mechanics with a higher contribution at stance and end of swing
4	Hip ABD/ADD	Moment before 25% stride	Hip abduction moment at single support

Abbreviations:
INV = inversion; EVE = eversion; ABD = abduction; ADD = adduction; PF = plantarflexion; DF = dorsiflexion; FLEX = flexion; EXT = extension; IR = internal rotation; ER = external rotation; % VAF = percentage variance accounted for

References

R Core Team (2017), R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL https://www.R-project.org/.

Arellano, C.J., Kram, R., 2014. Partitioning the metabolic cost of human running: A task-by-task approach. Integr Comp Biol 54, 1084-98.

Benson, L.C., Cobb, S.C., Hyngstrom, A.S., Keenan, K.G., Luo, J., O'Connor, K.M., 2018. Identifying trippers and non-trippers based on knee kinematics during obstacle-free walking. Hum Mov Sci 62, 58-66.

Boccia, G., Zoppirolli, C., Bortolan, L., Schena, F., Pellegrini, B., 2018. Shared and task-specific muscle synergies of Nordic walking and conventional walking. Scand J Med Sci Sports 28, 905-18.

Bohm, H., Hosl, M., Schwameder, H., Doderlein, L., 2014. Stiff-knee gait in cerebral palsy: how do patients adapt to uneven ground? Gait Posture 39, 1028-33.

Chou, L.-S., Draganich, L.F., 1998. Placing the trailing foot closer to an obstacle reduces flexion of the hip, knee, and ankle to increase the risk of tripping. J Biomech 31, 685-91.

Clark, D.J., Ting, L.H., Zajac, F.E., Neptune, R.R., Kautz, S.A., 2010. Merging of healthy motor modules predicts reduced locomotor performance and muscle coordination complexity post-stroke. J Neurophysiol 103, 844-57.

Cole, G.K., Nigg, B.M., Ronsky, J.L., Yeadon, M.R., 1993. Application of the joint coordinate system to three-dimensional joint attitude and movement representation: A standardization proposal. J Biomech Eng 115, 344-9.

Fukuchi, C.A., Fukuchi, R.K., Duarte, M., 2018. A public dataset of overground and treadmill walking kinematics and kinetics in healthy individuals. PeerJ 6, e4640.

Fukuchi, R.K., Fukuchi, C.A., Duarte, M., 2017. A public dataset of running biomechanics and the effects of running speed on lower extremity kinematics and kinetics. PeerJ 5, e3298.

Gottschall, J.S., Kram, R., 2005. Energy cost and muscular activity required for leg swing during walking. J Appl Physiol (1985) 99, 23-30.

Harshman, R.A., 1970. Foundations of the PARAFAC procedure: Models and conditions for an ‘‘explanatory’’ multimodal factor analysis. UCLA Working Papers in Phonetics 16, 1–84.

Harshman, R.A., 1972. PARAFAC2: Mathematical and technical notes. UCLA Working Papers in Phonetics 22, 30-44.

Harshman, R.A., Lundy, M.E., 1994. PARAFAC: Parallel factor analysis. Computational Statistics & Data Analysis 18, 39-72.

Helwig, N.E., 2018, multiway: Component Models for Multi-Way Data. R package version 1.0-5. [https://CRAN.R-project.org/package=multiway].

Helwig, N.E., Hong, S., Bokhari, E., 2013. Analyzing individual and group differences in multijoint multiwaveform gait data using the Parafac2 model. Int J Numer Method Biomed Eng 29, 62-82.

Helwig, N.E., Hong, S., Polk, J.D., 2012. Parallel Factor Analysis of gait waveform data: A multimode extension of Principal Component Analysis. Hum Mov Sci 31, 630-48.

Kristianslund, E., Krosshaug, T., van den Bogert, A.J., 2012. Effect of low pass filtering on joint moments from inverse dynamics: Implications for injury prevention. J Biomech 45, 666-71.

Leardini, A., Sawacha, Z., Paolini, G., Ingrosso, S., Nativo, R., Benedetti, M.G., 2007. A new anatomically based protocol for gait analysis in children. Gait Posture 26, 560-71.

Liew, B., Helwig, N.E., Morris, S., Netto, K., 2018. Influence of proximal trunk borne load on lower limb countermovement joint dynamics J Biomech [Epub ahead of print].

Oliveira, A.S., Gizzi, L., Ketabi, S., Farina, D., Kersting, U.G., 2016. Modular Control of Treadmill vs Overground Running. PLoS ONE 11, e0153307.

Saito, A., Tomita, A., Ando, R., Watanabe, K., Akima, H., 2018. Muscle synergies are consistent across level and uphill treadmill running. Sci Rep 8, 5979.

Schache, A.G., Baker, R., 2007. On the expression of joint moments during gait. Gait Posture 25, 440-52.

Schwartz, M.H., Rozumalski, A., 2008. The gait deviation index: A new comprehensive index of gait pathology. Gait Posture 28, 351-7.

Srivastava, S., Kao, P.C., Reisman, D.S., Higginson, J.S., Scholz, J.P., 2016. Coordination of muscles to control the footpath during over-ground walking in neurologically intact individuals and stroke survivors. Exp Brain Res 234, 1903-14.

Steele, K.M., Rozumalski, A., Schwartz, M.H., 2015. Muscle synergies and complexity of neuromuscular control during gait in cerebral palsy. Dev Med Child Neurol 57, 1176-82.

Van Campenhout, A., Bar-On, L., Aertbelien, E., Huenaerts, C., Molenaers, G., Desloovere, K., 2014. Can we unmask features of spasticity during gait in children with cerebral palsy by increasing their walking velocity? Gait Posture 39, 953-7.

van den Hoorn, W., Hodges, P.W., van Dieen, J.H., Hug, F., 2015. Effect of acute noxious stimulation to the leg or back on muscle synergies during walking. J Neurophysiol 113, 244-54.