



OPEN

Non-negative matrix factorisation is the most appropriate method for extraction of muscle synergies in walking and running

Mohammad Fazle Rabbi^{1,2}, Claudio Pizzolato^{1,2}✉, David G. Lloyd^{1,2}, Chris P. Carty^{1,2,3,4}, Daniel Devaprakash^{1,2} & Laura E. Diamond^{1,2}

Muscle synergies provide a simple description of a complex motor control mechanism. Synergies are extracted from muscle activation patterns using factorisation methods. Despite the availability of several factorisation methods in the literature, the most appropriate method for muscle synergy extraction is currently unknown. In this study, we compared four muscle synergy extraction methods: non-negative matrix factorisation, principal component analysis, independent component analysis, and factor analysis. Probability distribution of muscle activation patterns were compared with the probability distribution of synergy excitation primitives obtained from the four factorisation methods. Muscle synergies extracted using non-negative matrix factorisation best matched the probability distribution of muscle activation patterns across different walking and running speeds. Non-negative matrix factorisation also best tracked changes in muscle activation patterns compared to the other factorisation methods. Our results suggest that non-negative matrix factorisation is the best factorisation method for identifying muscle synergies in dynamic tasks with different levels of muscle contraction.

Despite decades of research, it remains unclear how the human central nervous system (CNS) coordinates activity of a large number of muscles during movement. Numerous studies suggest that the CNS activates muscles in groups to reduce the complexity required to control each individual muscle when performing a movement^{1,2}. The coordination of muscles activated in synchrony is commonly referred to as muscle synergy³. Indirect evidence suggests that muscle synergies reside in the brainstem and/or spinal cord and follow a modular organization^{4–6}. Muscle synergies are regarded as low dimensional units that produce complex activation patterns for a group of muscles, typically recorded via electromyography (EMG), during performance of a task^{4,6}. Synergies can be observed at cortical⁷ or spinal⁸ levels, suggesting a high degree of cooperation within the CNS's structural hierarchy⁹. Understanding the organisation of muscle synergies may help elucidate the neurological mechanisms that underpin a multitude of neurological conditions, including stroke^{10–12}, cerebral palsy^{13,14}, spinal cord injury¹⁵, and Parkinson's Disease¹⁶.

Factorisation methods use recorded and processed EMG signals, from here referred to as muscle activation patterns, to quantify muscle synergies. A number of different factorisation methods have been used to extract muscle synergies from muscle activation patterns during dynamic tasks. The four most commonly used factorisation methods reported in the literature between 1999–2018 are non-negative matrix factorisation (NMF)^{13,14,17–20} (62.28%), principal component analysis (PCA)^{21–24} (23.11%), independent component analysis (ICA)^{25,26} (3.22%), and factor analysis (FA)^{27,28} (2.15%) (Literature search in Supplementary material). Applying a factorisation method to a set of muscle activation patterns yields two components: (i) synergy weights; and (ii) synergy excitation primitives. The synergy weights scale the neural command intensity of the excitation primitives to represent the time varying neural command²⁹. Each factorisation method makes a different assumption

¹School of Allied Health Sciences, Griffith University, Gold Coast, QLD 4222, Australia. ²Griffith Centre of Biomedical and Rehabilitation Engineering (GCORE), Menzies Health Institute Queensland, Gold Coast, QLD 4222, Australia.

³Department of Orthopaedic Surgery, Children's Health Queensland Hospital and Health Service, Brisbane, QLD 4101, Australia. ⁴Research Development Unit, Caboolture and Kilcoy Hospitals, Metro North Hospital and Health Service, Brisbane, QLD 4101, Australia. ✉e-mail: c.pizzolato@griffith.edu.au

regarding the variance of input data and employs a different algorithm to extract the muscle synergies. The NMF algorithm is implemented using the multiplicative update rule based on Euclidian distance objective function¹⁷, gradient descent³⁰ and least square³¹ methods, and can be applied to both Gaussian and non-Gaussian datasets. NMF constrains muscle synergies to be non-negative and utilises second order statistics to find vectors that best describe the data's variance. PCA extracts the muscle synergies that best describe the data's variance while minimising covariance of the basis vectors (i.e. muscle synergy weights), and works best with Gaussian distributed datasets²⁹. The basis vector minimisation is solved analytically using singular value decomposition, returning the eigenvectors of the data's covariance matrix as the PCA identified muscle synergy weights. ICA is designed to deal with non-Gaussian variation in datasets and finds basis vectors (i.e. muscle synergy weights) that maximise the absolute value of the fourth moment of the data (i.e. kurtosis). Kurtosis is a measure of tailed-ness of the probability distribution of a real-valued random variable³² (indicating non-Gaussian data). Using pre-whitening, ICA transforms the data orthogonally to make it statistically independent^{29,33}. Apart from these methods, other algorithms also implement ICA, including Infomax ICA (maximizes Shannon mutual information)³⁴ and Kernel ICA (optimize variance to implement kernel Hilbert space)³⁵. Similar to PCA, FA employs eigenvalue decomposition to produce eigenvectors (i.e. muscle synergy weights) of the covariance matrix. Extracted muscle synergies with eigenvalues >1 are considered significant and all synergies with eigenvalues <1 are considered noise³⁶. All four factorisation methods use second or higher order statistics to estimate the muscle synergies from muscle activation patterns.

The performance of a factorisation method is highly dependent on variance in the recorded EMG signals. Variance accounted for (VAF) is typically employed to quantify reconstruction accuracy³⁰, whereby muscle activation patterns are reconstructed by linearly combining synergy weights and excitation primitives. VAF statistically compares experimental and reconstructed muscle activation patterns, but does not provide any insight regarding the magnitude of neural information conveyed by the recorded EMG signal. Alternatively, the probability distribution of a signal, quantified by the probability density function (PDF), provides EMG spectral information and has a direct relationship with the neural and peripheral information conveyed by the signal itself^{37,38}. Since different factorisation methods make different assumptions regarding the distribution of the muscle activation patterns, the probability distribution of the recorded EMG signal may indeed influence factorisation performance³⁰. Previous research has demonstrated that the probability distribution of a recorded EMG signal can be characterised by a Gaussian during high and medium intensity isometric muscle contractions, and by a super Gaussian (Laplacian) function during low intensity isometric muscle contractions^{32,39,40}. However, the probability distribution of recorded EMG signals from lower limb muscles during dynamic tasks such as walking and running is yet to be investigated.

NMF is the most commonly used factorisation method for muscle synergy analysis (Literature search in Supplementary material), perhaps because it is readily available in many commercial data processing packages (e.g. MATLAB). However, there is no clear evidence to support the use of NMF in place of other factorisation methods. Previous studies that compared factorisation methods primarily considered synthetic^{30,41}, animal (e.g., frog³⁰, cat²³), and upper limb²⁹ EMG datasets. Notably, Ivanenko and colleagues identified lower limb muscle synergies in humans during various walking conditions which remained stable when different factorisation methods were used⁴². However, none of these aforementioned studies compared factorisation methods for extraction of lower limb muscle synergies across both walking and running conditions. Furthermore, the probability distribution of included muscle activation patterns, an essential assumption of every factorisation method, has not previously been utilised as a metric for comparison. Therefore, this study aimed to use a PDF-based metric, along with VAF, to compare performance of the four most commonly used factorisation methods (NMF, PCA, ICA, FA) for extraction of lower limb muscle synergies during walking and running at different speeds. Given that NMF is the only factorisation method that makes no assumptions regarding distribution of the muscle activation patterns (i.e. Gaussian, non-Gaussian), we hypothesised that muscle synergies extracted using NMF would best capture the variation in PDF of muscle activation patterns during walking and running. Accordingly, we anticipated that muscle synergies extracted using NMF would best capture the variation of muscle activity due to gait speed and thereby better reconstruct the muscle activation patterns compared to the other factorisation methods.

Methods

Eighteen healthy participants (72% male, age = 25.8 ± 6.1 years, height = 1.75 ± 0.76 m, mass = 64.9 ± 8.2 kg, body mass index = 21.1 ± 1.9 kg/m²) with no recurrent or recent lower limb injuries participated. This study was approved by the relevant institutional Human Research Ethics Committee of Griffith University (GU Ref no. 2017/020) and Australian Institute of Sport (Ref no. 20170801) and, all participants provided written informed consent prior to participation in the study. All methods were carried out in accordance with relevant guidelines and regulations of Helsinki.

Data collection and processing. Surface EMG signals were recorded at 1500 Hz (Noraxon, Scottsdale, AZ, USA) from ten lower limb muscles: medial gastrocnemius, lateral gastrocnemius, soleus, peroneus longus, peroneus brevis, tibialis anterior, vastus lateralis, rectus femoris, biceps femoris, and sartorius. Skin preparation and electrode placement were performed according to SENIAM guidelines⁴³. Participants completed ten trials of over-ground walking (1.3 ± 0.1 m/s) and running at slow (3 ± 0.3 m/s), moderate (5 ± 0.5 m/s), and fast (7 ± 0.7 m/s) speeds. Timing gates were used to ensure that average speed for each trial remained within 10% of the condition's predefined limit.

Raw EMG signals were band-pass filtered at 30–400 Hz, full wave rectified, and low-pass filtered at 6 Hz using a zero-lag 4th order Butterworth filter to generate EMG linear envelopes. For each trial (consisting of one full cycle, toe off to toe off), the linear envelopes were amplitude normalised to the trial's peak value⁴⁴ and time normalised to

	Walk		Slow run		Moderate run		Fast run	
	Raw EMG	Muscle activation patterns	Raw EMG	Muscle activation patterns	Raw EMG	Muscle activation patterns	Raw EMG	Muscle activation patterns
Kurtosis	1.75(±0.19)	1.72(±0.20)	1.45(±0.21)	1.42(±0.20)	1.29(±0.19)	1.25(±0.18)	1.1(±0.2)	1.0(±0.10)
Skewness	0.01(±0.02)	0.012(±0.02)	0.009(±0.02)	0.012(±0.01)	0.005(±0.01)	0.01(±0.01)	0.003(±0.01)	0.009(±0.01)

Table 1. Average (\pm standard deviation) probability density function kurtosis and skewness of raw EMG signals and muscle activation patterns from 10 muscles for 19 participants during walking and running at different speeds (10 trials each). FA: factor analysis; ICA: independent component analysis; NMF: non-negative matrix factorisation; PCA: principal component analysis.

200 data points⁴⁵, to ensure data from each muscle, trial, and task were equally represented in the muscle synergy analysis. Therefore, the processed EMG dataset (i.e. muscle activation patterns) computed from each trial consisted of one matrix with 10 rows (1 per muscle) and 200 columns (1 per time point).

Muscle synergy analysis. Muscle activation patterns (x) can be reconstructed by multiplying the muscle synergy weights (w) and synergy excitation primitives (h):

$$x(t) = \sum_{i=1}^N w_i h_i(t) + e(t) \quad (1)$$

where e represents reconstruction error. Equation (1) can be written in vector form as

$$x = wh + e \quad (2)$$

where x is a $L \times N$ matrix representing the muscle activation patterns of L muscles at N time points, w is a $L \times K$ matrix containing time independent muscle weightings with K being the number of muscle synergies, and h is a $K \times N$ matrix of time varying synergy excitation primitives, and e is a $L \times N$ matrix representing the reconstruction error of each muscle's activation pattern at each time point^{46,47}. Reconstructed muscle activation patterns can be compared with experimental muscle activation patterns to evaluate the performance of the factorisation method used to identify the muscle synergies. In all four factorisation methods, muscle activation patterns are factorised by minimising the mean square error (MSE) between original and reconstructed muscle activation patterns. In general, the objective function can be written as $\min \|x - \tilde{x}\|$. In NMF $\tilde{x} = wh$ and $w_i \geq 0; h_i \geq 0$ is assumed for non-negativity⁴⁸. In FA and PCA, $\tilde{x} = wh^T$ with $h^T h = I$ assumed in PCA⁴⁹. In ICA, $\tilde{x} = Awh$ where A represents a mixing matrix which is used for pre-whitening of data^{50,51}. Equation (2) can be re-written as

$$x = \tilde{x} + e \quad (3)$$

To evaluate the variation in probability distribution of muscle activity at different walking/running speeds, we calculated the PDF of both the recorded (raw) EMG signals and the muscle activation patterns. Since the factorisation method is only applied to the muscle activation patterns and not to the raw EMG, the PDF of the muscle activation patterns was compared with the PDF of the excitation primitives. The average PDF for all recorded muscles followed a transition from a super Gaussian to Gaussian nature (Table 1). Thus, the reconstruction error can also be assumed Gaussian-distributed and variance of the error (σ_e^2) can be used to calculate MSE as

$$MSE = var(\tilde{x}) + \sigma_e^2 \quad (4)$$

If the distribution of x and h is the same (given the other parameter - muscle synergy weights (w) - is independent), then the error variance (σ_e^2) is known. Also, given that w represents scalar weights that linearly scale and combine h to produce \tilde{x} , then the PDF of \tilde{x} is given by PDF of h . In other words, the best estimate of muscle activation (\tilde{x}) can be calculated when x and h have the same distribution. Thus, the similarity between probability distribution of x and h is a metric that can be used to compare factorisation methods.

Comparison of factorisation methods. The PDF of the muscle activation patterns was modelled using a Gaussian probability distribution function

$$P(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \quad (5)$$

where μ_x , and σ_x represent the mean and standard deviation of muscle activation patterns, respectively. Similarly, the PDF of the synergy excitation primitives identified by the factorisation methods was calculated as

$$P(h) = \frac{1}{\sigma_h \sqrt{2\pi}} e^{-\frac{(h-\mu_h)^2}{2\sigma_h^2}} \quad (6)$$

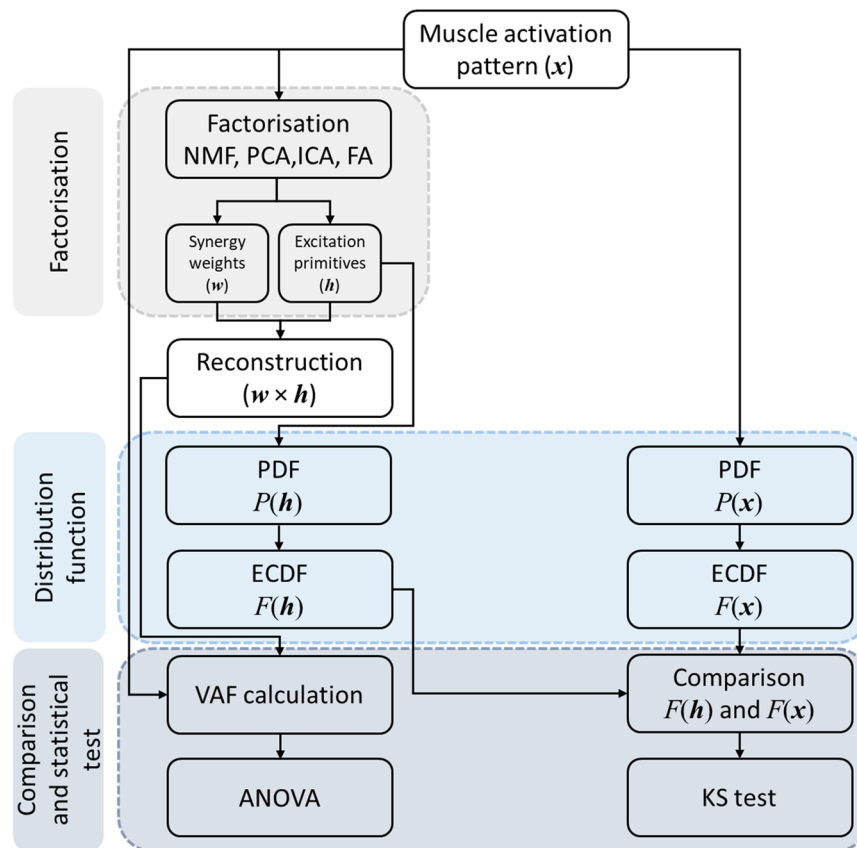


Figure 1. Workflow describing the methods used to compare the performance of four factorisation methods for muscle synergy analysis.

where μ_h , and σ_h represent the mean and standard deviation of excitation primitives, respectively.

The Kolmogorov-Smirnov (KS) test was used to compare the probability distribution of muscle activation patterns and synergy excitation primitives. The KS test is based on differences in both location and shape of the cumulative distribution functions of two samples regardless of the type of distributions and the number of samples⁵². The cumulative distributions of $P(x)$ and $P(h)$ was calculated using

$$F(x) = \int_{-\infty}^x P(x)dx \quad (7)$$

$$F(h) = \int_{-\infty}^x P(h)dh \quad (8)$$

The cumulative distributions represented by Eqs. (7) and (8) are equivalent to the empirical cumulative distribution function (ECDF)⁵³, which estimates the cumulative distribution of a random variable by assigning equal probability to each observation in a sample. The ECDF allows for two distributions to be compared without knowing the exact probability distributions associated to x and h . The KS test yields the maximum dissimilarity between the ECDF of x and h , where the dissimilarity between $F(x)$ and $F(h)$ was calculated

$$D = \max|F(x) - F(h)| \quad (9)$$

As the number of muscle synergies is always less than the number of muscles (i.e. $K < L$), there is a dimension mismatch between $F(x)$ and $F(h)$ in Eq. (9). To compare these two ECDFs, we assessed the similarity between the ECDF of each k^{th} synergy excitation primitive and all L muscle activation patterns by calculating maximum dissimilarities between two ECDFs as

$$D_k = \max|F_l(x) - F_k(h)|, l = 1, 2, \dots, L; k = 1, 2, \dots, K \quad (10)$$

After calculating D_k using equation (10), the maximum dissimilarity index (D) was calculated as the maximum of D_k among all k as $D = \max|D_k|$. Maximum dissimilarity represents the single worst-case scenario of similarity between the ECDFs of the excitation primitives and muscle activation patterns. This workflow (Fig. 1) was repeated for all four factorisation methods. If the ECDFs of x and h were not hugely different, the KS test failed to reject the null hypothesis (i.e. significant differences between ECDFs occurred at $p < 0.05$)^{54,55}. In

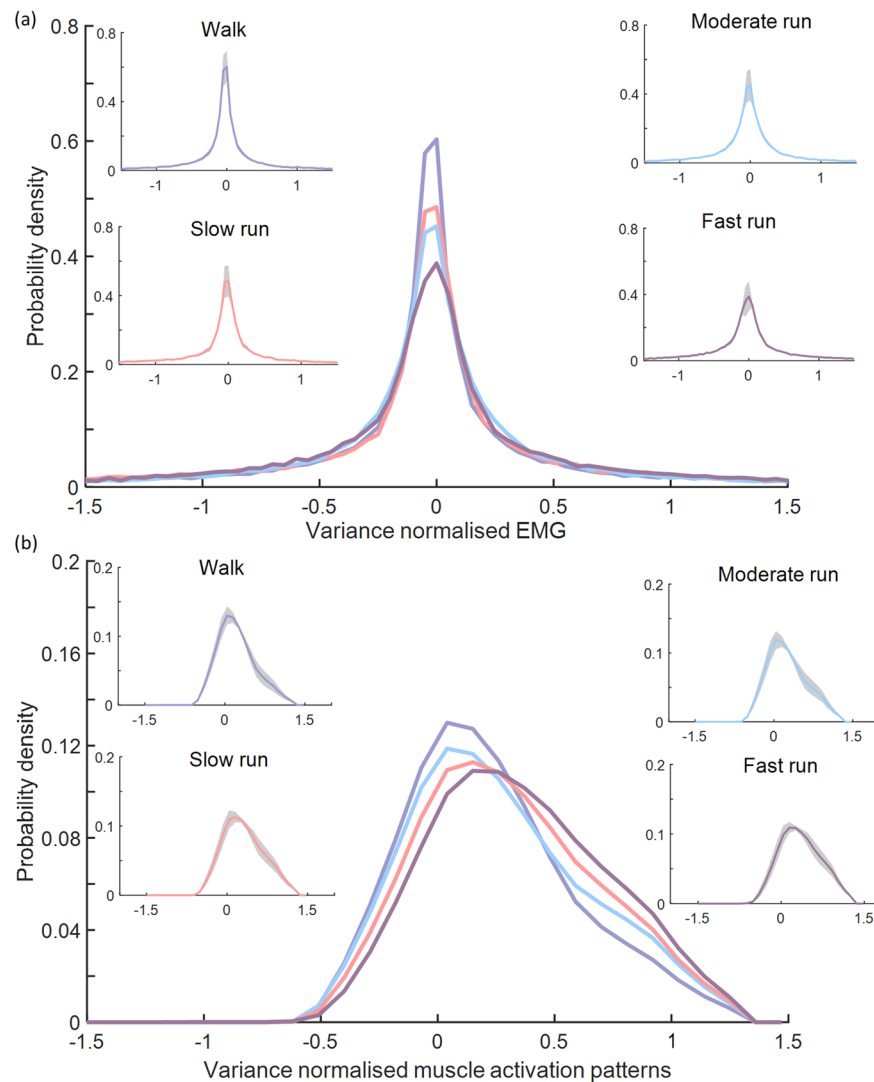


Figure 2. Average probability distribution of (a) raw electromyogram (EMG) signals and (b) muscle activation patterns from 10 lower limb muscles (medial gastrocnemius, lateral gastrocnemius, soleus, peroneus longus, peroneus brevis, tibialis anterior, vastus lateralis, rectus femoris, biceps femoris, and sartorius) from a representative participant during walking and slow, moderate, and fast running conditions. Shaded regions indicate standard deviation across 10 trials. The probability density function for both the raw EMG and muscle activation patterns widened as a function of increasing gait speed.

addition to the maximum dissimilarity, percentage of occurrence of agreement (i.e. rate of failure to reject the null hypothesis at $p < 0.05$) was also used to evaluate the similarity between the probability distribution of muscle activation patterns and synergy excitation primitives.

The reconstruction performance of each factorisation method varies according to the number of muscle synergies used³⁰. VAF can be calculated to determine the minimum number of muscle synergies needed to adequately reconstruct muscle activation patterns for each task

$$VAF = 1 - \frac{\|(\mathbf{x} - \mathbf{w} * \mathbf{h})^2\|}{\|\mathbf{x}\|^2} \quad (11)$$

VAF was calculated for a number of synergies (K) between 2 and 6 for each of the four selected factorisation methods. A final value of K was selected with VAF greater than 85% for all factorisation methods. The minimum number of synergies was then used to compare factorisation methods using the KS test.

The factorisation methods were compared based on VAF between the original and reconstructed muscle activation patterns and PDF similarity was compared between the original muscle activation patterns and synergy excitation primitives (Fig. 1). The PDF similarity was assessed using occurrence of agreement and maximum dissimilarity index for each gait speed. The individual metrics for each factorisation method across all gait speeds were compared using a repeated measure analysis of variance (ANOVA) with Bonferroni correction.

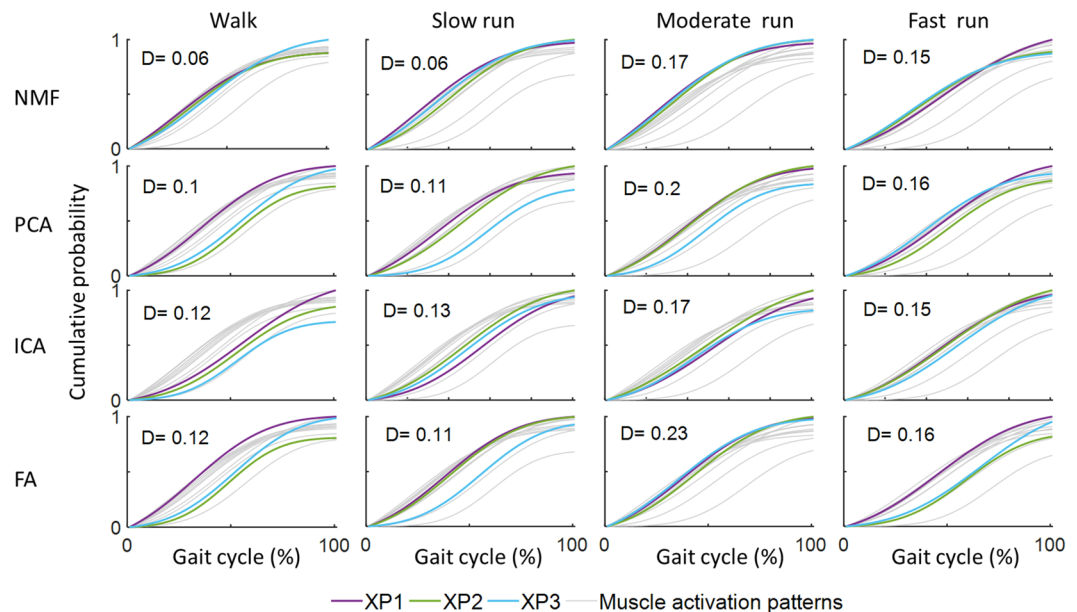


Figure 3. ECDF between muscle activation patterns and synergy excitation primitives, and corresponding maximum dissimilarities (D) (equation 10) for one representative participant. Lower maximum dissimilarity indicates greater similarity between ECDFs from the muscle activation patterns and synergy excitation primitives.

Results

The mean PDF of the raw EMG signals and muscle activation patterns from 10 muscles from a representative participant are shown in Fig. 2, while the PDF and ECDF of raw EMG signals from individual muscles in all participants are reported in Supplementary Figs. S1 and S2, respectively. Average kurtosis of the PDFs across all participants decreased as gait speed increased, indicating that the PDF is transformed from Laplacian (theoretical kurtosis = 3)^{32,56} to Gaussian (theoretical kurtosis = 0)^{32,56} with increasing speed (Table 1). Minimal positive skewness of the PDFs was observed as gait speed increased, indicating high symmetry between the left and right side of the PDF⁴⁰.

Extracting three muscle synergies ($k=3$) accounted for >85% of variance in muscle activation patterns for all factorisation methods. Typically, NMF generated ECDFs with greater similarity and lower maximum dissimilarities indices (D) between muscle activation patterns and synergy excitation primitives, compared to the other factorisation methods (Fig. 3).

VAF was significantly higher using NMF than the other factorisation methods ($p < 0.01$) for walking. It was also higher than PCA and ICA ($p < 0.05$) for the slow run condition, and higher than PCA ($p < 0.01$) for the moderate and fast run conditions (Fig. 4a). NMF in general resulted in the highest occurrence of agreement of ECDF for all gait speed conditions (Fig. 4b), and was found to be significantly higher than PCA and ICA for walking ($p < 0.01$). PCA and FA had higher agreement than ICA for the slow and moderate run conditions ($p < 0.01$). In addition, NMF had the lowest dissimilarity among all factorisation methods for all speed conditions (Fig. 4c). The maximum dissimilarity of NMF was found to be significantly lower than ICA for walking and for slow and moderate run conditions ($p < 0.01$), and lower than FA for slow, moderate ($p < 0.05$), and fast ($p < 0.01$) run conditions. An example of reconstruction of muscle activation patterns using four factorisation methods for four gait speeds is reported in Fig. 5. Original and reconstructed muscle activation patterns of individual muscles from all participants for four gait speeds are shown in Supplementary Fig. S3.

Discussion

This study aimed to use a PDF-based metric, along with VAF, to compare performance of the four most commonly used factorisation methods (NMF, PCA, ICA, FA) for extraction of lower limb muscle synergies during walking and running at different speeds. In agreement with our hypotheses, NMF demonstrated superior performance for all three metrics when three muscle synergies were extracted. These findings suggest that NMF may have unique ability to capture changes in the probability distribution of muscle activation patterns across different speeds, indicating that NMF is the best factorisation method for identifying muscle synergies in dynamic tasks with different levels of muscle contraction.

Our analysis identified that the PDF of the raw EMG signals from 10 lower limb muscles widened as a function of increasing gait speed. Notably, the probability distribution transitioned from Laplacian to Gaussian as gait speed increased, which is supported by a corresponding decrease in kurtosis and skewness. This transition is in line with previous investigations where similar changes in PDF of upper-limb EMG signals were observed with increasing isometric contraction intensity^{39,40}, and in agreement with animal studies^{57,58}. Given that the factorisation methods were applied to the muscle activation patterns and not directly to the raw EMG signals, we verified

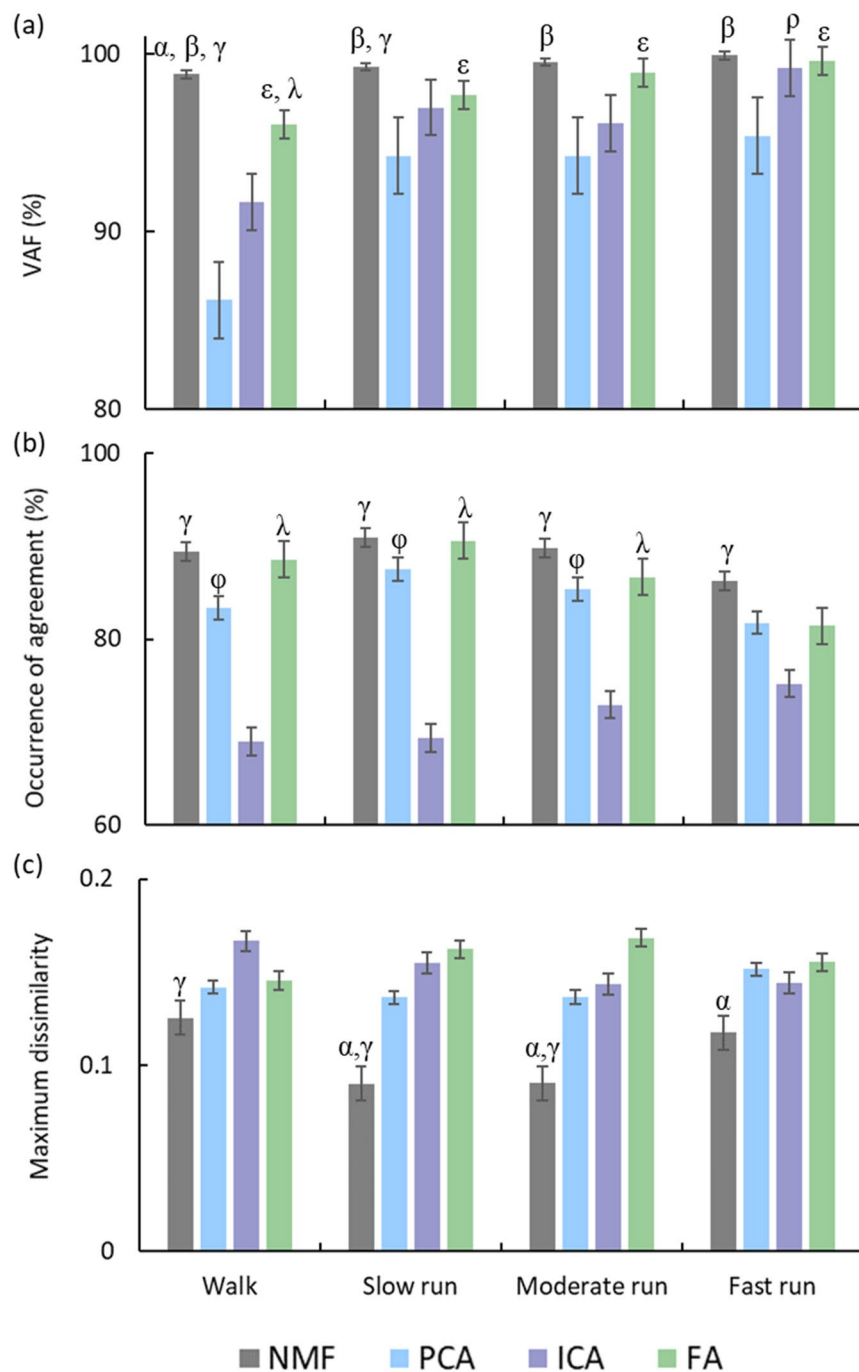


Figure 4. (a) VAF (%), (b) occurrence of agreement (%), and (c) maximum dissimilarity between the empirical cumulative distribution function (ECDF) of the muscle activation patterns and synergy excitation primitives at four gait speeds. Error bars indicate standard deviation. The following symbols represent significant differences ($p < 0.05$) between factorisation methods, α : NMF > FA or NMF < FA, β : NMF > PCA, γ : NMF > ICA or NMF < ICA, ϵ : FA > PCA, λ : FA > ICA, ρ : ICA > PCA, φ : PCA > ICA.

that the PDF of the muscle activation patterns demonstrated similar behaviour to the PDF of the raw EMG signals (Table 1; Fig. 2). As both PDFs vary as a function of gait speed, it is essential that a factorisation method capture these variations in order to reconstruct muscle activation patterns effectively. We examined the variation of PDF and ECDF of excitation primitives across four gait speeds (Supplementary Figs. S4 and S5). PDFs of the excitation primitives from all factorisation methods vary as a function of speed, except those obtained from FA (Supplementary Table T1). We also assessed whether the PDF of the muscle activation pattern was affected by the low-pass filter cut-off frequency used in creating the linear envelop. We evaluated the PDF using three different cut-off frequencies (10 Hz, 8 Hz and 6 Hz). Results demonstrate that the PDF of individual muscle activation patterns remain unaffected by cut-off frequency in this range (Supplementary Fig. S6).

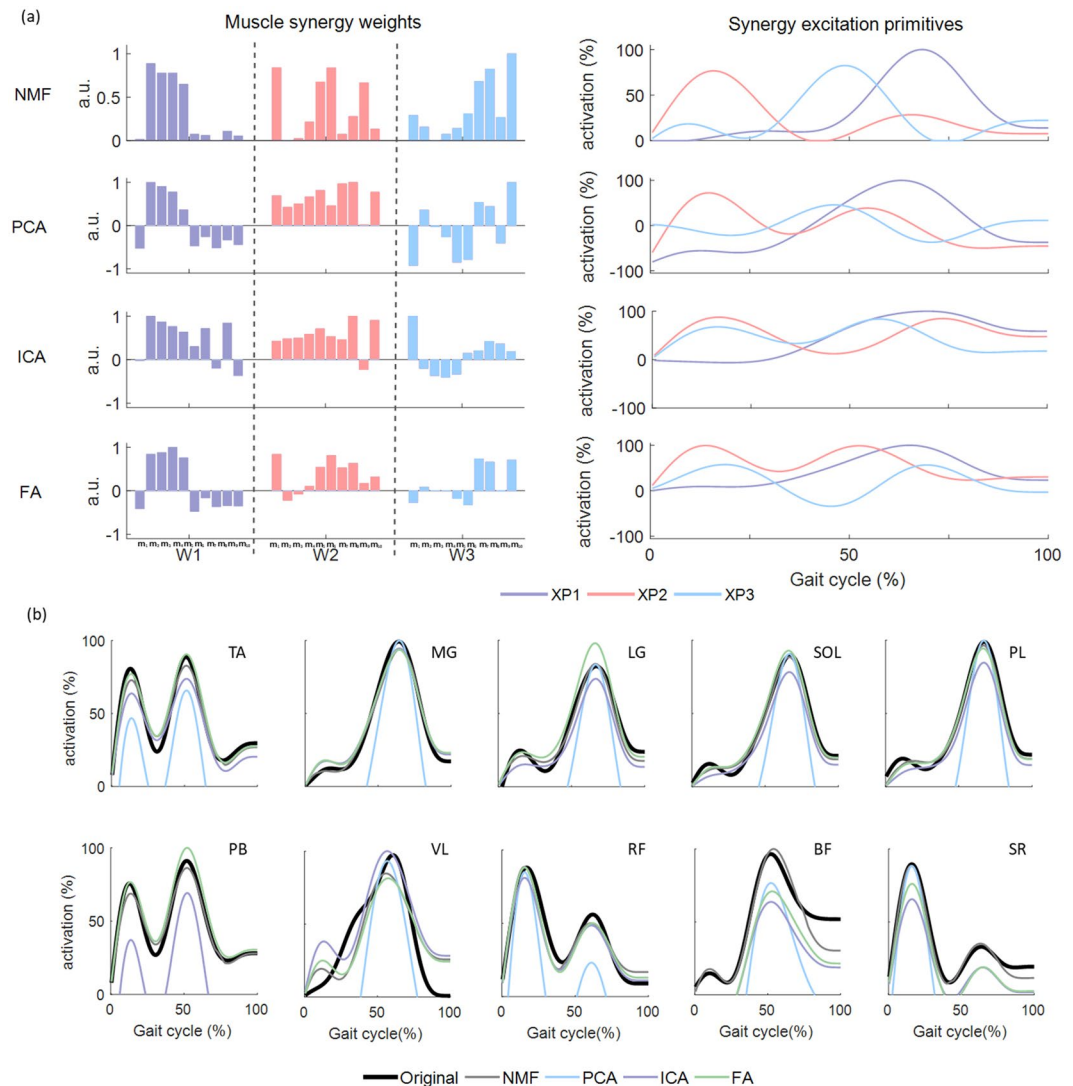


Figure 5. Reconstruction of muscle activation patterns with three muscle synergies using each factorisation method during walking for one representative participant: **(a)** muscle synergy weights (W_1 - W_3) from 10 included muscles and synergy excitation primitives identified using NMF, PCA, ICA and FA, **(b)** original muscle activation patterns and reconstructed muscle activation patterns for 10 included muscles. a.u.: arbitrary unit.; m_1 : biceps femoris; m_2 : lateral gastrocnemius; m_3 : medial gastrocnemius; m_4 : peroneus longus; m_5 : peroneus brevis; m_6 : rectus femoris; m_7 : sartorius; m_8 : soleus; m_9 : tibialis anterior; m_{10} : vastus lateralis.

Three metrics (VAF, occurrence of agreement, and maximum dissimilarity) were used to compare performance of NMF, PCA, ICA, and FA using three muscle synergies (Fig. 4). NMF performed significantly better than the other factorisation methods for at least one metric for each gait speed condition. When considering VAF alone, the factorisation methods performed equally for running; however, when PDF similarity was also taken into account, NMF outperformed the other factorisation methods. When we repeated our analysis using four muscle synergies, our primary metrics remained relatively unchanged (Supplementary Fig. S7). VAF is a statistical measure of signal reconstruction accuracy, while occurrence of agreement and maximum dissimilarity are used to compare the PDF of muscle activation patterns and synergy excitation primitives. Maximum dissimilarity was used as a goodness of fit to find the most similarity between the ECDFs of the excitation primitives and muscle activation patterns. Maximum dissimilarity will result in a high value even if the means of the two ECDFs are the same but the shape of the ECDFs are different (e.g. the tail of one ECDF is longer)⁵⁹. Unlike VAF, the probability distribution (as assessed by occurrence of agreement and maximum dissimilarity) has a direct relationship with peripheral and neural information conveyed by the recorded EMG signal during any muscle activity^{37,60}. Given the unique yet complimentary assessment parameters of VAF and PDF, implementing both metrics is likely to provide a more comprehensive understanding of reconstruction performance. Though we only considered walking and running at different speeds, it seems plausible that these results could be extrapolated to different dynamic tasks of varying intensities.

Our results demonstrate that NMF can capture the variation in PDF associated to speed-related changes in the raw EMG signal. As a direct consequence of this unique capability, the muscle synergies extracted using NMF provided the best representation of the muscle activation patterns recorded during gait at different speeds. In a previous study, muscle synergies extracted via NMF were shown to be similar across different gait (both walking and running) speeds when compared the muscle synergy weights⁴⁵. As such, it may be possible to use only a single set of NMF identified muscle synergies to reconstruct the continuum of gait speeds, at least in healthy individuals. However, it is unclear whether our observations would be reproduced when investigating neurologically-impaired cohorts or different dynamic tasks.

This study has some limitations that warrant consideration. First, only healthy individuals were included and therefore these findings may not be generalisable to other populations. Nonetheless, since NMF does not rely on the type of distribution (Gaussian, non-Gaussian) of muscle activation patterns, it is likely to be the most sensitive method for characterising neuro-related alterations in the PDF. We utilised three muscle synergies (based on >85% VAF) for all four factorisation methods, which is in line with previous investigations using NMF^{20,61}, PCA²³, ICA²⁹ and FA²⁷ for walking and running at different speeds. Further, our primary metrics (VAF, occurrence of agreement, and maximum dissimilarity) remained relatively unchanged when four (Supplementary Fig. S7), five, six, seven, and eight synergies were extracted (data not presented).

NMF demonstrated superior performance for extraction of lower limb muscle synergies during walking and running at different speeds compared to PCA, ICA, and FA, based partly on our implementation of a PDF-based metric which identified that NMF can uniquely evolve with changes in gait speed. As such, NMF should be preferred for muscle synergy analysis applications that involve gait with variable speeds. Future investigations are required to confirm these findings in other dynamic tasks with varying intensities.

Data availability

The datasets analysed in the current study is available from the corresponding author on responsible request.

Received: 11 November 2019; Accepted: 28 April 2020;

Published online: 19 May 2020

References

- d'Avella, A., Saltiel, P. & Bizzi, E. Combinations of muscle synergies in the construction of a natural motor behavior. *Nat Neurosci* **6**, 300–308 (2003).
- Ting, L. H. & McKay, J. L. Neuromechanics of muscle synergies for posture and movement. *Current Opinion in Neurobiology* **17**, 622–628 (2007).
- Ferrante, S. *et al.* A personalized multi-channel FES controller based on muscle synergies to support gait rehabilitation after stroke. *Frontiers in Neuroscience* **10** (2016).
- Tresch, M. C., Saltiel, P., d'Avella, A. & Bizzi, E. Coordination and localization in spinal motor systems. *Brain Res Brain Res Rev* **40**, 66–79 (2002).
- Roh, J., Cheung, V. C. K. & Bizzi, E. Modules in the brain stem and spinal cord underlying motor behaviors. *Journal of Neurophysiology* **106**, 1363–1378 (2011).
- Bizzi, E. & Cheung, V. C. The neural origin of muscle synergies. *Frontiers in Computational Neuroscience* **7**, 1–6 (2013).
- Overduin, S. A., d'Avella, A., Roh, J., Carmenta, J. M. & Bizzi, E. Representation of muscle synergies in the primate brain. *Journal of Neuroscience* **35**, 12615–12624 (2015).
- Ivanenko, Y. P., Poppele, R. E. & Lacquaniti, F. Spinal cord maps of spatiotemporal alpha-motoneuron activation in humans walking at different speeds. *Journal of neurophysiology*, (2006).
- Santuz, A. *et al.* Modular control of human movement during running: an open access data set. *Frontiers in physiology* **9**, (2018).
- Clark, D. J., Ting, L. H., Zajac, F. E., Neptune, R. R. & Kautz, S. A. Merging of Healthy Motor Modules Predicts Reduced Locomotor Performance and Muscle Coordination Complexity Post-Stroke. *Journal of Neurophysiology* **103**, 844–857 (2010).
- McGowan, C. P., Neptune, R. R., Clark, D. J. & Kautz, S. A. Modular control of human walking: Adaptations to altered mechanical demands. *Journal of Biomechanics* **43**, 412–419 (2010).
- Cheung, V. C. *et al.* Muscle synergy patterns as physiological markers of motor cortical damage. *Proc Natl Acad Sci USA* **109**, 14652–14656 (2012).
- Steele, K. M., Rozumalski, A. & Schwartz, M. H. Muscle synergies and complexity of neuromuscular control during gait in cerebral palsy. *Dev Med Child Neurol* **57**, 1176–1182 (2015).
- Schwartz, M. H., Rozumalski, A. & Steele, K. M. Dynamic motor control is associated with treatment outcomes for children with cerebral palsy. *Developmental Medicine and Child Neurology* **58**, 1139–1145 (2016).
- Pérez-Nombela, S. *et al.* Modular control of gait after incomplete spinal cord injury: differences between sides. *Spinal Cord* **55**, 79–86 (2017).
- Rodriguez, K. L., Roemmich, R. T., Cam, B., Fregly, B. J. & Hass, C. J. Persons with Parkinson's disease exhibit decreased neuromuscular complexity during gait. *Clinical Neurophysiology* **124**, 1390–1397 (2013).
- Lee, D. D. & Seung, H. S. Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature* **401**, 788–791 (1999).
- Rozumalski, A., Steele, K. M. & Schwartz, M. H. Muscle synergies are similar when typically developing children walk on a treadmill at different speeds and slopes. *Journal of Biomechanics* **64**, 112–119 (2017).
- Shuman, B. R. *et al.* Repeatability of muscle synergies within and between days for typically developing children and children with cerebral palsy. *Gait & Posture* **45**, 127–132 (2016).
- Saito, A., Tomita, A., Ando, R., Watanabe, K. & Akima, H. Muscle synergies are consistent across level and uphill treadmill running. *Scientific reports* **8**, 5979–5979 (2018).
- Ting, L. H. & Macpherson, J. M. A limited set of muscle synergies for force control during a postural task. *Journal of neurophysiology* **93**, 609–613 (2005).
- Ting, L. H. *et al.* Neuromechanical principles underlying movement modularity and their implications for rehabilitation. *Neuron* **86**, 38–54 (2015).
- Ting, L. H. & Chvatal, S. A. In *Motor Control: Theories, Experiments, and Applications* 102–138 (Oxford University Press, 2010).
- Falaki, A., Huang, X., Lewis, M. M. & Latash, M. L. Motor equivalence and structure of variance: multi-muscle postural synergies in Parkinson's disease. *Experimental Brain Research* **235**, 2243–2258 (2017).
- Hyvarinen, A. & Oja, E. Independent Component Analysis: Algorithms and Applications. *Neural Networks* **13**, 411–430 (2000).

26. Hart, C. B. & Giszter, S. F. Distinguishing synchronous and time-varying synergies using point process interval statistics: motor primitives in frog and rat. *Frontiers in Computational Neuroscience* **7**, 1–14 (2013).
27. Kiehlba, P. *et al.* How are Muscle Synergies Affected by Electromyography Pre-Processing? *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26**, 882–893 (2018).
28. Saito, A., Watanabe, K. & Akima, H. Coordination among thigh muscles including the vastus intermedius and adductor magnus at different cycling intensities. *Human Movement Science* **40**, 14–23 (2015).
29. Lambert-Shirzad, N. & Van der Loos, H. F. M. On identifying kinematic and muscle synergies: a comparison of matrix factorization methods using experimental data from the healthy population. *Journal of Neurophysiology* **117**, 290–302 (2017).
30. Tresch, M. C., Cheung, V. C. K. & d'Avella, A. Matrix Factorization Algorithms for the Identification of Muscle Synergies: Evaluation on Simulated and Experimental Data Sets. *Journal of Neurophysiology* **95**, 2199–2212 (2006).
31. Chih-Jen, L. Projected Gradient Methods for Non-negative Matrix Factorization. *Neural computation* **19**, 2756–2779 (2007).
32. Thongpanja, S. *et al.* Probability Density Functions of Stationary Surface EMG Signals in Noisy Environments. *IEEE Transactions on Instrumentation and Measurements* **65**, 1547–1557 (2016).
33. Suzuki, T. & Sugiyama, M. Least-Squares Independent Component Analysis. *Neural Computation* **23**, 284–301 (2011).
34. Bell, A. J. & Sejnowski, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural computation* **7**, 1129–1159 (1995).
35. Bach, F. R. & Jordan, M. I. Kernel independent component analysis. *Journal of machine learning research* **3**, 1–48 (2002).
36. Davis, B. L. & Vaughan, C. L. Phasic behavior of EMG signals during gait: use of multivariate statistics. *Journal of Electromyography and Kinesiology* **3**, 51–60 (1993).
37. Farina, D. *et al.* The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **22**, 797–809 (2014).
38. Kaufman, M., Zurcher, U. & Sung, P. S. Entropy of electromyography time series. *Physica A: Statistical Mechanics and its Applications* **386**, 698–707 (2007).
39. Clancy, E. A. & Hogan, N. Probability density of the surface electromyogram and its relation to amplitude detectors. *IEEE Transactions on Biomedical Engineering* **46**, 730–739 (1999).
40. Nazarpour, K., Al-Timemy, A. H., Bugmann, G. & Jackson, A. A note on the probability distribution function of the surface electromyogram signal. *Brain Research Bulletin* **90**, 88–91 (2013).
41. Ebied, A., Kinney-lang, E., Spyrou, L. & Escudero, J. Evaluation of matrix factorisation approaches for muscle synergy extraction. *Medical Engineering and Physics* **0**, 1–10 (2018).
42. Ivanenko, Y. P., Cappellini, G., Dominici, N., Poppele, R. E. & Lacquaniti, F. Coordination of locomotion with voluntary movements in humans. *Journal of Neuroscience* **25**, 7238–7253 (2005).
43. Hermens, H. J. *et al.* European recommendations for surface electromyography. *Roessingh research and development* **8**, 13–54 (1999).
44. Devaprakash, D., Weir, G. J., Dunne, J. J., Alderson, J. A. & Donnelly, C. J. The influence of digital filter type, amplitude normalisation method, and co-contraction algorithm on clinically relevant surface electromyography data during clinical movement assessments. *Journal of Electromyography and Kinesiology* **31**, 126–135 (2016).
45. Cappellini, G., Ivanenko, Y. P., Poppele, R. E. & Lacquaniti, F. Motor patterns in human walking and running. *Journal of neurophysiology* **95**, 3426–3437 (2006).
46. Santuz, A., Ekizos, A., Janshen, L., Baltzopoulos, V. & Arampatzis, A. On the Methodological Implications of Extracting Muscle Synergies from Human Locomotion. *International Journal of Neural Systems* **27**, 1750007–1750007 (2017).
47. Sartori, M., Gizzi, L., Lloyd, D. G. & Farina, D. A musculoskeletal model of human locomotion driven by a low dimensional set of impulsive excitation primitives. *Front Comput Neurosci* **7**, 79 (2013).
48. Lee, D. D. & Seung, H. S. In Advances in neural information processing systems. 556–562.
49. Udell, M., Horn, C., Zadeh, R. & Boyd, S. Generalized low rank models. *Foundations and Trends® in Machine Learning* **9**, 1–118 (2016).
50. Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks* **10**, 626–634 (1999).
51. Hendrikse, A., Veldhuis, R. & Spreeuwiers, L. In Proceedings of the 28th Symposium on Information Theory in the Benelux, Enschede, The Netherlands. 211–218 (Eindhoven).
52. Massey, F. J. Jr The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association* **46**, 68–78 (1951).
53. Van der Vaart, A. W. Asymptotic statistics. Vol. 3 (Cambridge university press, 2000).
54. Dressler, M. *et al.* A potential mechanism for age-related declines in patellar tendon biomechanics. *Journal of Orthopaedic Research* **20**, 1315–1322 (2002).
55. Hernandez, F., Shull, P. B. & Camarillo, D. B. Evaluation of a laboratory model of human head impact biomechanics. *Journal of biomechanics* **48**, 3469–3477 (2015).
56. Westfall, P. H. Kurtosis as peakedness, 1905–2014. RIP. *The American Statistician* **68**, 191–195 (2014).
57. Hutchison, D., Roy, R., Hodgson, J. & Edgerton, V. EMG amplitude relationships between the rat soleus and medial gastrocnemius during various motor tasks. *Brain research* **502**, 233–244 (1989).
58. Hutchison, D. L., Roy, R. R., Bodine-Fowler, S., Hodgson, J. A. & Edgerton, V. R. Electromyographic (EMG) amplitude patterns in the proximal and distal compartments of the cat semitendinosus during various motor tasks. *Brain research* **479**, 56–64 (1989).
59. Sridharan, R. In Statistics for Research Projects- MIT Course No. IAP 2015 (MIT, Cambridge MA, 2019).
60. Farina, D., Merletti, R. & Enoka, R. M. The extraction of neural strategies from the surface EMG: an update. *Journal of Applied Physiology* **117**, 1215–1230 (2014).
61. Oliveira, A. S., Gizzi, L., Farina, D. & Kersting, U. G. Motor modules of human locomotion: influence of EMG averaging, concatenation, and number of step cycles. *Frontiers in Human Neuroscience* **8** (2014).

Acknowledgements

This work was funded by a Griffith University postgraduate research scholarship (MFR), Australian Research Council (LP150100905) and Menzies Health Institute, Queensland, Australia. The authors would like to thank Dr David J. Saxby, Professor Rod Barrett, Dr Bryce Killen, Mr. Giorgio Davico for their contribution to the project and all participants for their support.

Author contributions

Conceptualisation, M.F.R., C.P., D.G.L.; Methodology, M.F.R., C.P., D.G.L., C.P.C., D.D. and L.E.D.; Investigation, M.F.R., C.P., C.P.C., D.D. and L.E.D.; Formal Analysis, M.F.R., C.P., C.P.C. and L.E.D.; Writing – Original Draft, M.F.R. and C.P.; Writing – Review & Editing, M.F.R., C.P., D.G.L., C.P.C., D.D. and L.E.D.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-65257-w>.

Correspondence and requests for materials should be addressed to C.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020