Unbiased estimation of the population-level motor module

Yusuke Matsui[1,2,*], Kohei Uno[1], Ippei Nojima[3]

[1]Biomedical and Health Informatics Unit, Department of Integrated Health Science, Nagoya University Graduate School of Medicine, 461-8673 Nagoya, Aichi, Japan

[2]Institute for Glyco-core Research (iGCORE), Nagoya University, 461-8673 Nagoya, Aichi, Japan

[3]Graduate School of Medical Sciences, Department of Physical Therapy, Nagoya City University

*Corresponding author:

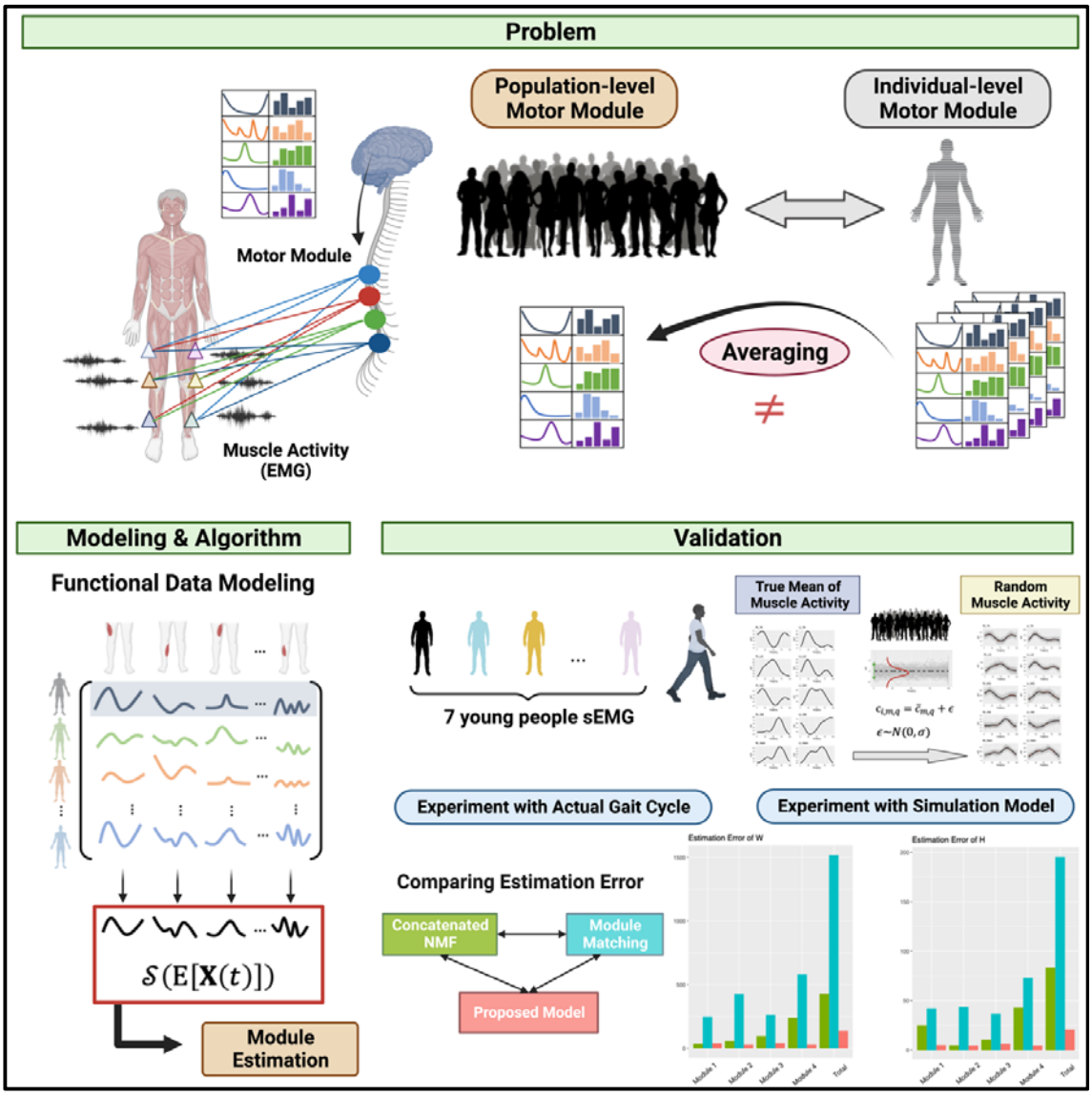
Yusuke Matsui

Summary

Motor module is a functional neurophysiological command for muscle coordination. In clinical settings, population-level characterization and comparison of motor modules are necessary to evaluate pathophysiological mechanisms and intervention effects. Previous studies have estimated individual motor modules and then compared them, but the validity of capturing the distribution of the latent population has not been fully understood. Our study aimed to address this issue by investigating the accuracy of estimating the population mean of motor modules. Through simulation experiments, we found that previous individual-based approach did not converge regardless of sample size and was vulnerable to noise. We developed an unbiased estimation algorithm using the framework of functional data analysis, which significantly improved estimation accuracy. Our findings highlight statistical challenges for motor module analysis and suggest the need for further research on new computational algorithms using large-scale clinical data.

Keywords: Motor module, computational approach, muscle activity, functional data analysis, estimation, model

Graphical Abstract



Introduction

Motor control is a complex process that involves coordinated commands for thousands of motor units in hundreds of skeletal muscles, generated by the central nervous system in real-time to produce movements. To help explain this complexity, the concept of a motor module or muscle synergy has been proposed[1–6]. A motor module is a set of functional neurophysiological commands that drive muscle coordination[1,7]. Currently, motor modules are characterized via dimension-reduction techniques such as non-negative matrix factorization (NMF) [8–10](**Figure 1A**). However, NMF only finds two matrices that minimize residual errors for each dataset, and the resulting values have no biological meaning[11,12]. This makes it challenging to compare or summarize motor modules across multiple subjects (**Figure 1C**). Previous computational methodological study has mainly focused on developing sophisticated models for individual motor modules and improving their estimation reproducibility[13–15]. However, to gain a better understanding of the generality of motor modules, it is also crucial to establish their structure at the population level. Unfortunately, however, statistical and computational methods for this purpose have not yet been fully elucidated.

To address these challenges, two main approaches have been emerged. One approach involves module matching between subjects[16–20], where individual motor modules are estimated and then matched to each other. However, this approach is not always guaranteed

to give a one-to-one solution, and the repeated application of NMF individually can contaminate motor module estimates by local solutions due to initial value dependence. Another approach is joint learning called concatenated NMF (cNMF), where subjects' data are combined into one pseudo-data as input[11,21,22].This approach can help to avoid some of the problems with module matching, but it can also be numerically unstable and lead to convergence issues as the matrix size increases with the number of subjects. Overall, developing reliable statistical and computational methods for characterizing motor modules at the population level is an ongoing challenge, but it is crucial to advancing our understanding of motor control and its organization across individuals. To address these issues, this study presents an unbiased estimation algorithm for motor modules at the population level using NMF and Functional Data Analysis (FDA) [23,24] (**Figure 1D**). Our approach accounts for population variability and reduces statistical errors in motor module expectation estimation, particularly in larger sample sizes where other methods fail to converge.

Results

Population-level motor module estimation

In this section, we statistically formulated the population-level motor module borrowing the concept of FDA[23,24]. We decompose the variability in muscle activity into two levels: within-individual variability and between-individual variability. Within-individual variability accounts for variations within a single subject, such as trial-to-trial variation in the

gait cycle, which can be reduced through preprocessing steps like filtering and temporal alignment. In contrast, between-individual variability arises from differences between subjects. We use the FDA to model this variability, taking into account the population characteristics. In the discussion that follows, individual muscle activity data shall be subjected to appropriate preprocessing, such as filtering, outlier removal, temporal alignment, and summarization, for inter-subject comparison.

To describe between-subjects variability in muscle activity, individual muscle activity patterns are described by functional random variables. Let $X_m(t)$ be a random functional variable, where the activity of skeletal muscle m at time t is represented by the function. The muscle activity $x_{i,m}(t)$ of subject i that we observe can be viewed as the observed values sampled from the functional population Ω. The observed muscle activity of individual subject can be denoted with the population mean

$$\mu_m(t) = \mathrm{E}[X_m(t)]$$

and the error term $\epsilon_m(t)$ as

$$x_{i,m}(t) = \mu_m(t) + \epsilon_{i,m}(t) \#(1)$$

where $\epsilon_m(t)$ is parameterized by $\mathrm{E}[\epsilon_m(t)] = 0$ and $\mathrm{Var}[\epsilon_m(t)] = \sigma$, representing population-level variability. The typical distribution that $X_m(t)$ follows is the normal distribution, but if non-negative constraints are considered, a non-negative exponential family of distribution (such as the Poisson distribution) can also be assumed[25]. In practice, to deal

with digitized discrete data, time t is discretized as a vector t with length T:

$$x_{i,t,m} = \mu_{t,m} + \epsilon_{i,t,m} \#(2)$$

Our proposed population-level motor module can be estimated using $\mu_{t,m}$. The population mean of muscle activity, expressed in matrix form with $\mu_{t,m}$ as an element, is

$$\mathbf{U} = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,M} \\ \mu_{2,1} & \mu_{2,2} & \cdots & \mu_{2,M} \\ \vdots & \vdots & \cdots & \vdots \\ \mu_{T,1} & \mu_{T,2} & \cdots & \mu_{T,M} \end{bmatrix} \#(3)$$

and motor module can be estimated by dimension reduction algorithm. In case of NMF, the mean muscle activity matrix $\mathbf{U}$ can be decomposed as

$$\mathbf{U} = \mathbf{WH} + \mathbf{E} \#(4)$$

where $\mathbf{E}$ is the estimation error term matrix. The resulting matrix $\mathbf{W}$ and $\mathbf{H}$ are population version of primitive signals and synergy, respectively.

Existing approaches

Next, here we compare the proposed method with two existing approaches to population mean estimation of motor modules: the module matching approach and cNMF approach. The major difference between the proposed method and the two existing approaches is that the population mean is estimated after first estimating each individual motor module. Denote by $\mathbf{X}$ the matrix whose elements are $x_{i,t,m}$:

$$\mathbf{X}_i = \begin{bmatrix} x_{i,1,1} & x_{i,1,2} & \cdots & x_{i,1,M} \\ x_{i,2,1} & x_{i,2,2} & \cdots & x_{i,2,M} \\ \vdots & \vdots & \cdots & \vdots \\ x_{i,T,1} & x_{i,T,2} & \cdots & x_{i,T,M} \end{bmatrix} \#(5)$$

For the module matching approach, individual motor modules are first estimated:

$$\mathbf{X}_i = \mathbf{W}_i\mathbf{H}_i + \mathbf{E} \#(6)$$

In contrast to the proposed method, the order of the synthesized variables (i.e., rows of **W** and columns of **H**) is arbitrary, so matching process is critical to make the modules correspond across subjects. One way to find the module of subject j that corresponds to the synergy k of subject i is to find a module of subject j that maximizes the correlation coefficient. That is,

$$k' = \text{argmax}_l \, cor\left(\mathbf{w}_{ik}, \mathbf{w}_{jl}\right) \ \#(7)$$

where $\mathbf{w}_{ik}$ and $\mathbf{w}_{jl}$ are column vector of matrix $\mathbf{W}_i$ and $\mathbf{W}_j$, respectively. To estimate the population mean of the motor module estimation, the mean of the matrix **W** or **H** is calculated for each subject after matching.

In this approach, the final **W** and **H** include mismatch errors due to the fact that the optimal number of modules (i.e., rows of **W** and columns of **H**) can be different across the subjects and there is no guarantee of a one-to-one correspondence between modules, and estimation errors resulting from local solutions due to initial value dependency of NMF algorithm[26,28].

On the other hand, cNMF also performs individual module estimation, but it estimates individual motor modules by combining subjects as one pseudo-subject, applying NMF, and re-splitting the obtained results for each subject. To construct the input matrix, the

individual muscle activity matrices $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}$ are concatenated row by row

$$\mathbf{X}^{c} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \#(8)$$

then, estimate the motor module:

$$\mathbf{X}^{c} = \mathbf{W}^{[c]}\mathbf{H}^{[c]}$$

Splitting $\mathbf{W}^{[c]}$ into the original sample-wise blocks for $1,2,\ldots,n$ corresponds to the primitive signal for each subject. Then, the population mean of the motor module estimation can be estimated by averaging these estimates. The cNMF approach assumes that the synergy $\mathbf{H}^{[c]}$ is common across samples[11], which is implicitly considered equivalent to estimating the population mean $\mathbf{H}$ across subjects. Thus, cNMF can be otherwise described as a method of estimating $\mathbf{W}^{[c]}$ conditioned on the population mean $\mathbf{H}$, which is expected to give results close to those of the proposed method.

The cNMF algorithm naturally results in larger matrices with increasing sample sizes, the use of NMF on these matrices can present challenges due to the higher number of parameters that must be estimated. First, the possibility of initial value-dependent local optimum solutions is further exacerbated. Also, the inclusion of some outliers can affect the process of sequential optimization, making it less robust to noise and increasing the likelihood of bias in the estimated results. It also increases the computational cost in terms of memory size and iteration to convergence.

Evaluation with an actual data set

The proposed method was examined using the muscle activity during the gait cycle in seven young healthy subjects (**Figure 3A**). Based on the preprocessed muscle activities, the expected motor modules were estimated using three methods: the proposed method, module matching approach, and cNMF-based approach (**Figure 3B**). The number of motor modules was determined by the variance accounted for (VAF) (**Star Methods**). We selected 4 as the number of motor modules based on a $\mathrm{VAF} \geq 99\%$.

Although both methods showed similar trends for both **W** and **H**, inter-method variability in the estimates was evident (**Figure 3B**). For example, module 4 showed a marked difference in the timing of the increase in the primitive signal (**Figure 3B**). This could lead to substantially conflicting conclusions regarding the timing of CNS activity during gait cycles between the proposed and other methods. There was also significant variation between the methods in module estimation (**Figure 3B**), with the individual-based approach indicating that the entire muscle was involved in any module, whereas the cNMF-based approach and the proposed method suggested a sparse group structure. There were also significant differences between the cNMF-based approach and the proposed method, for example, in modules 3 and 4, indicating differences in the inference to the motor module during gait cycles (**Figure 3B**).

Simulation model for evaluating methods

Next we evaluated the degree to which the expected value of the motor module could be reproduced for numerically generated population-level muscle activity data (**Figure 4A**). Equation (1) allows us to represent individual muscle activity patterns as samples from a probability distribution with a population mean function and variance. To obtain sample-wise muscle activity that follows a given probability distributions, we first decompose population expectation of muscle activity into several linear combinations based on basis functions (e.g., B-splines functions). The mean function $\mu_m(t)$ can be decomposed as

$$\mu_m(t) = \sum_{q=1}^{Q} \bar{c}_{m,q} b_{m,q}(t) \#(9)$$

A set of mutually orthogonal functions $b_{m,q}(t)$ are used to decompose the population mean function $\mu_m(t)$ into Q functional linear combinations, each of which has weight coefficients $\bar{c}_{m,q}$. Here the coefficient vector $\bar{\mathbf{c}}_m = \left(\bar{c}_{m,1}, \bar{c}_{m,2}, \ldots, \bar{c}_{m,Q}\right)$ has sufficient information to reconstruct the population mean function under the basis function. In the FDA framework, the variation of muscle activity among subjects in a population can be described by variation with respect to this coefficient vector. In the numerical experiments, a normal distribution with mean average $\bar{c}_{m,q}$ and variance σ was assumed for individual muscle activity.

$$\mathrm{c}_{i,m,q} = \bar{c}_{m,q} + \epsilon \#(10)$$

$$\epsilon \sim N(0, \sigma)$$

168 Substituting the sampled coefficient vectors back into equation (9) yields the individual

169 muscle activities that follow the population distribution.

$$x_{i,m}(t) = \sum_{q=1}^{Q} c_{i,m,q} b_{m,q}(t) \#(11)$$

170 The FDA-based simulation model can be used to evaluate whether the expected

171 values estimated from individual motor module estimates reproduce the expected values of

172 the motor modules derived from the population mean of muscle activity. The residual sum of

173 squares (RSS) of the true motor module was evaluated to assess the performance of each

174 method. In other words, for the RSS of k −th module, we evaluated the

$$\mathrm{RSS}_k^{[W]} = \sum_{t=1}^{T} \left(w_{tk}^{[true]} - w_{tk}^{[estimated]} \right)^2 \#(12)$$

$$\mathrm{RSS}_k^{[H]} = \sum_{m=1}^{M} \left(h_{km}^{[true]} - h_{km}^{[estimated]} \right)^2 \#(13)$$

175 where $w_{tk}^{[estimated]}$ and $h_{km}^{[\mathrm{estimated}]}$ are estimated values for $\mathbf{W}^{[true]}$ and $\mathbf{H}^{[true]}$,

176 respectively.

177 **The proposed model allows for control of estimation bias**

178 We used a simulation model to assess each method (**Figure 4A**) with generated muscle

179 activity for n=10, 50, and 100 (**Figure 4B**). While all methods captured the primitive signal's

180 trend, there was a bias in estimated motor modules based on individual and cNMF compared

181 to the true motor module (**Figure 4C**). The individual approaches showed a tendency to

182 overestimate the module weights i.e., averaged $\mathbf{H}_i$ in equation (6) over subjects after

matching (**Figure 4C**). Our proposed method had accuracy close to that of cNMF, but with less estimation bias (**Figure 4D**). The fourth module showed an overestimation of the weight of the right leg TA (n=50) and underestimation of the left leg LG with cNMF (n=100) compared to the proposed method, with differences in the timing of peak signals (**Figure 4C, 4D**). We assessed the estimation bias of each method by calculating the RSS residuals for incremental changes in sample size. Our results showed that the proposed method converges to an estimation bias of zero as the sample size increases, in contrast to the other methods (**Figure 5B**). This suggests that existing methods lack asymptotic properties when estimating the population mean of motor modules.

Evaluating noise robustness

To assess the impact of deviating values on estimation accuracy, we conducted a similar evaluation as in the previous section with 10% of the samples containing deviating signal values (**Figure 6A, 6B**). For small samples (n=10), all methods showed large estimation errors (**Figure 6C, 6D**). However, as the sample size increased (n=50, n=100), the proposed method significantly reduced the error compared to the other methods (**Figure 7A**). When we also examined the convergence of the estimation error with increasing sample size, the estimation bias of the proposed method converged to zero, while the other methods showed no tendency to converge (**Figure 7B**). Specifically, the estimation bias of cNMF increased

significantly compared to the case without deviating values (**Figures 5A, 5B; Figures 7A, 7B**), indicating its sensitivity to such values.

Discussion

We proposed an algorithm for estimating the motor module at the population level using Non-Negative Matrix Factorization (NMF) within the framework of Functional Data Analysis (FDA). Our method first estimates the population mean of muscle activity and then performs motor module estimation based on it. In contrast, the existing approach estimates the population mean after estimating the motor module for individual. The difference between the two approaches may seem simple, but our approach has been shown to significantly reduce estimation bias beyond a certain sample size and is also robust to noise.

Although our proposed algorithm showed convergence to the true motor module patterns with only 25 to 30 samples, the existing approaches failed as the sample size increased (**Figure 5B**). This was true even in the presence of samples that contained some deviant signal values (**Figure 6B**). Various factors can contribute to the failure of existing methods to converge. For example, the module-matching method can cause forced matching errors when there is no corresponding module, while cNMF may suffer from numerical instability or bias contamination due to collinearity or deviation values as the number of parameters to be optimized increases with the number of samples.

Our results indicate that our proposed algorithm outperforms existing methods in terms of noise robustness and convergence to true motor module estimates. However, we also found that the presented algorithm had a rather large estimation bias in small-sample conditions, and full convergence required larger samples (**Figure6B-D, Figure 7A, 7B**). Therefore, while our proposed algorithm is an improvement over existing methods, its application may be limited in situations where the variability of the population is high and the sample size is very small. These results highlighted the need for statistical improvements in computational algorithms.

Alternative methods, such as principal component analysis (PCA) [27,29,30] and factor analysis (FA) [31,32], can also be utilized for motor module estimation, but in this study, we focused on NMF. One of the reasons for this choice is that NMF has been frequently employed and shown promise in estimation accuracy for various tasks, such as walking and running[13]. However, all these methods share common issues related to dimensionality reduction, such as the matching of synthetic variables' order and sign, and potential contamination from heterogeneous observations, such as outliers[33,34]. Therefore, it is essential to further investigate and improve upon the accuracy and estimation methods for population-level motor module parameter estimation based on dimensionality reduction in the future direction.

This study has several limitations. Firstly, the validity of the estimation algorithm needs to be comprehensively evaluated for other commonly used movement tasks besides periodic walking. Secondly, the actual data analysis was performed using a small number of samples, and further evaluation with a larger clinical dataset is necessary. Thirdly, the presented algorithm assumes a homogeneous distribution of the latent population, and when multiple heterogeneous subpopulations are mixed, subpopulations must be identified and stratified. In this regard, future tasks include identifying subgroups based on dimension reduction[35] and clustering[36,37] and developing methods to test for heterogeneity based on functional data representation.

Characterizing motor modules at the population level is a crucial step towards translating experimental findings into clinical applications. For example, evaluation of intervention effects on motor function before and after intervention[38–49], and large cross-sectional clinical studies[38,50,51] need to capture effects as a population. The present study is the first to investigate the motor module estimation algorithm from a population-level perspective. Although the study has its limitations, such as the small sample size, it provides valuable insights into the estimation of motor modules at the population level. Continued development of the algorithms will fully realize their application to a broader range of movement tasks and clinical situations.

Acknowledgments

We would like to thank Editage (http://www.editage.jp) for editing and reviewing this manuscript for English language.

Author Contributions

The study conceptualization was led by YM. Data analysis and interpretation were carried out by YM and IN. Algorithm development and evaluation were spearheaded by YM and KU. Manuscript drafting was a joint effort by YM and IN. All authors extensively reviewed, revised, and approved the final version of the manuscript for submission.

Declaration of Interests

The authors declare no competing interests.

Figure Legends

Figure 1. Conceptual view of this study

A) Description of the general motor module estimation algorithm. Multichannel Electromyography (EMG) data is given by a time × muscle matrix. Motor module analysis is a problem of unsupervised low-dimensional structural learning based on muscle activity data from multiple sites. Non-negative matrix factorization (NMF) is widely used as one of them. Each estimated dimension is considered to reflect the activity of the central nervous system, where **W** is called the primitive signal and is considered to be the input signal of the central nervous system, and **H** is called synergy and is considered to represent a potentially cooperating muscle group, or motor module. **B) Population-level motor module and individual-level module.** When conducting a multi-subject motor module analysis, there are two approaches depending on the purpose. The first is population-level motor module analysis, which is intended for characterization and inference at the population level and for comparison among populations. The second is individual-level motor module analysis, the main objectives of which are individual-level characterization and inter-individual comparisons. Our goal is to estimate population-level motor module expectations in an unbiased manner. We show that conventional methods are not consistent with true expectations where they estimate the expectation by aggregating individual-level motor modules. **C) Problems with conventional methods.** The main causes are 1) the arbitrariness

of the solution due to the unsupervised learning of NMF, 2) local minima problem with an initial value, 3) errors associated with integrating the results of motor module analysis from different subjects, and 4) instability of the numerical solution of the NMF algorithm. **D) Proposed method.** We introduce the framework of functional data analysis. Assuming a stochastic population of muscle activity, the unbiased motor module expectation is estimated based on the expectation of muscle activity.

Figure 2. Comparison of algorithms between existing and proposed methods

A) Module matching approach. Multichannel Electromyography (EMG) data given in a time × muscle matrix for each individual was used to estimate the motor module using NMF, and then they are matched and averaged across subjects. **B) Concatenated NMF-based approach.** The NMF is applied by concatenating the EMG matrices across subjects in the row direction. Since the columns of the muscle activity matrix are fixed, the estimated synergy **H** approximates the average across subjects, and the primitive signal **W** conditioned by the common synergy **H** is estimated for each subject. **C) Proposed method.** First, expectations of muscle activity are estimated from the EMG matrices of multiple subjects, and NMF is applied to the obtained expected values of muscle activity. Since the expected values of EMG reflect the population structure, the obtained motor modules also reflect the population structure.

Figure 3. Actual data example

A) Data overview and analysis design. The Multichannel Electromyography (EMG) signals of 10 left and right muscles and foot pressure were measured in seven healthy young adults. The data were processed using an in-house analysis pipeline that implements standard muscle activity preprocessing. These preprocessed data were used to compare results from existing and proposed methods. **B) Comparison of estimated motor modules.** Four modules were identified based on variance accounted for. The left panel represents module composition **H**, the columns represent each method, and each row represents motor modules 1 through 4. Although all modules represent close trends between methods, there is variation in the estimates and different interpretations of the muscle groups that constitute the motor module. The right panel represents the primitive signal **W** corresponding to the four synergies, which also shows a similar trend between methods, but some of the modules show different timing of activation, suggesting that the clinical interpretation, such as the correspondence with the gait phase, can be different.

Figure 4. Simulation experiments of estimation bias

A) Simulation model. The model was constructed to generate Multichannel Electromyography (EMG) based on functional data representation. True muscle activity μ_m in skeletal muscle m was approximated by linear combination by a small set of B-spline orthogonal basis functions $b_{m,q}$ by regression splines with coefficients $\bar{c}_{m,q}$. To mimic the simulation of real data, the mean muscle activity derived from the actual muscle activity of seven subjects was used. Assuming a normal distribution for the probabilistic distribution in the population, spline regression coefficients corresponding to individual subjects $c_{i,m,q}$ were generated by adding noise which follows a multivariate normal distribution to $\bar{c}_{m,q}$, and by multiplying the B-spline basis $b_{m,q}$ which was used for the true muscle activity approximation μ_m. Resulting individual muscle activity profiles follow a normal distribution in the sense of functional data representation. **B) Generated individual muscle activity and its expectation.** The individual muscle activity profiles generated for n=10, 50, and 100 samples (gray lines) and their expected values (red) and true muscle activity (black) are shown. **C) Comparison of estimation bias of motor module expectation.** We estimated the expectation of motor modules and compared the estimation bias with each method. Here, we defined “true” expectation of motor module (black) as one estimated from true muscle activity μ_m. The proposed method (red) gave an estimate close to the true motor module

348 expectation (black) in both primitive signal and modules. On the other hand, exiting methods

349 exhibited estimation bias in at least one of the modules.

350

Figure 5. Comparison of residual sum of squares for estimation biases

A) Accuracy of each method. The residual sum of squares (RSS) calculated based on the results shown in Figure 4 are displayed for each motor module. The rightmost RSS in each panel shows the sum of the RSS over all modules. In the simulation, the proposed method (red) has the lowest RSS for all modules. **B) Asymptotics of RSS with sample size.** The RSS for each module is shown when the sample size is increased by 5 from 5 to 100 samples. With the nature of statistical error, RSS should asymptotically approach zero as the sample size increases. Only the proposed method (red) showed asymptotic behavior in the simulations.

Figure 6. Simulation experiment of noise robustness

A) Simulation model. The simulation model shown in Figure 4 was modified to include some outliers. The muscle activity was generated by simulation assuming a population with 10% of outlier samples contaminated in addition to the normal sample. The generation of outlier samples was accomplished by adding errors with excess variance to a subset of the coefficients of the regression splines. The samples were then multiplied with the B-spline basis functions to obtain samples with distinct muscle activity profiles. **B) Generated individual muscle activity and its expectation.** The individual muscle activity profiles generated for n=10, 50, and 100 samples (gray lines) and their expected values (red) and true muscle activity (black) are shown. **C) Comparison of the estimation robustness of motor module expectation.** We estimated the expectation of the motor module and compared the robustness of estimation between methods. Under small-sample conditions, all methods showed estimation errors, but as the number of samples increased, the proposed method (red) gave estimates closer to the true motor module expectation.

Figure 7. Comparison of the residual sum of squares for robustness

A) Robustness of each method. The residual sum of squares (RSS) calculated based on the results shown in Figure 5 are displayed for each motor module. The rightmost RSS in each panel shows the sum of RSS over all modules. In the simulation, the proposed method (red) has the lowest RSS for all synergies, which indicates robustness to outliers. **B) Stability of RSS with sample size.** The RSS for each module is shown when the sample size is increased by 10 from 10 to 100 samples. It is expected that the effect of outliers will decrease as the sample size increases. However, only the proposed method (red) showed stability along with the sample size in this simulation.

STAR★Methods

Resource availability

Lead contact

matsui@met.nagoya-u.ac.jp

Materials availability

Data and code availability

The analysis code used in the paper can be accessed from the following repository URL (https://github.com/ymatts/RoMMS). The acquired data from this study are available from the lead contact upon reasonable request.

Method details

Data acquisition

Wireless electromyograms (EMGs) (Trigno EMG sensors, DELSYS, Boston, MA, United States) were recorded from the bilateral tibial anterior (TA), soleus (SOL), medial gastrocnemius (MG), lateral gastrocnemius (LG), rectus femoris (RF), vastus medialis (VM), medial hamstring (MH), and lateral hamstring (LH) muscles, according to SENIAM recommendations. The skin was gently abraded and cleaned with alcohol before the EMG recording. The sensors were placed as far as possible each other anatomically to minimize the potential risk of crosstalk between the EMG recordings. Data were collected for two minutes at a self-selected speed and used to examine steady-state gait by removing the first 10 s of the

gait. All the participants walked independently without walkers or crutches. The EMG signals were amplified (with a 909 gain preamplifier), band-pass filtered (10–450 Hz), and sampled at 1,000 Hz. Footswitches were attached to the heels to determine the time of foot contact.

Preprocessing

The foot pressure signal values acquired from the foot sensors were used to identify the gait cycle. The time point at which the signal value increased from zero was defined as the heel contact. sEMG signals were processed for each muscle. They were first low-pass filtered at 30 Hz using a fourth-order Butterworth filter and then rectified. Maximum voluntary contraction (MVC) normalization was performed at the maximum observation as 100%. An epoch was created for each gait cycle and identified using a foot sensor. To identify and exclude abnormal cycles, each trial was projected onto a two-dimensional space using a robust principal component analysis[52–54]. Using these coordinate values, trials that deviated significantly from the projected space were defined as outliers and were excluded. The threshold settings for the anomaly values followed the default values in the R package rospca[55]. The variation in time between the gait cycles was corrected according to linear length normalization (LLN), with 0% for the first heel contact and 100% for the next heel contact. Finally, a 10-Hz envelope was derived, and the average was calculated to obtain the muscle activity.

Computational algorithm

Because the parameter of time t is observed as a discrete value in the real world, the muscle activity $x_{i,m}(t)$ is expressed as $x_{i,t,m}$ using the discretized time $t = 1,2,\dots,T$. Equation (3) can be obtained using the time-discretized version of (4).

$$\bar{m} = \mathcal{S}(\mu_t) + \boldsymbol{\delta_s} \#(8)$$

where $\mu_t = \left(\mu_{t,1}, \mu_{t,2}, \dots, \mu_{t,M}\right)$ and

$$\mu_{t,m} = \frac{1}{n}\sum_{i=1}^{n} x_{i,t,m} \#(9)$$

The explicit form of the motor module by NMF in equation (6) can be defined using muscle activity matrix $\mathbf{U}^{[T\times M]} = \left\{\mu_{t,m}; t = 1,2,\dots,T, m = 1,2,\dots,M\right\}$. The primitive signals $\mathbf{W}^{[T\times K]}$ and synergy $\mathbf{H}^{[K\times M]}$ obtained by the NMF are derived as follows:

$$\mathbf{U} = \mathbf{WH} + \mathbf{E} \#(10)$$

where $\mathbf{E}$ is the estimation error term with the $T \times M$ matrix.

Analysis pipeline

The workflow of the analysis is shown in **Figure 3A**. The first step was to perform standard signal processing on the raw sEMG signal for each subject, including bandpass filtering, rectification, MVC normalization, envelope smoothing, and epoching. Then, time alignment was performed, if necessary. Finally, the trial average was calculated for each muscle to characterize the muscle activity pattern for one trial per subject. We then estimated the expected mean of the muscle activity in the subject population according to equation (9) and applied the NMF to obtain an estimate of the population mean of the motor module shown in

equation (10). For the number of modules, we used variance accounted for (VAF), defined below:

$$VAF_{K'} = 1 - \frac{RSS_{K'}}{\sum_{t=1}^{T}\sum_{m=1}^{M} x_{t,m}^2} \quad \#(11)$$

$$RSS_{K'} = \sum_{t=1}^{T}\sum_{m=1}^{M} \left(\hat{x}_{t,m} - x_{t,m}\right)^2$$

where $\hat{x}_{t,m}$ is reconstructed matrix of muscle activities using K' synergy.

Simulation model

This section presents a simulation model for comparing the performance of each method. Two simulations are conducted. The first is the accuracy with which each approach can reproduce the "true" motor modules in the population (**Figure 3A**). Here, "true" refers to the motor modules "estimated from the true population mean of muscle activity." Individual muscle activities were generated by adding normally distributed variations with a zero mean and a certain standard deviation to the predefined true population mean.

The second is the robustness of motor module estimation in the presence of outliers (**Figure 7A**). Assuming muscle activity from unrelated populations, the muscle activity distributed around different population means was prepared, and the muscle activity distributed around it was generated as an outlier sample. By mixing this with samples from the population of interest, we generated muscle activity that contained outliers and performed a simulation similar to the first to evaluate the reproducibility of the true motor modules.

460 The muscle activity patterns of each subject, which varied based on the population
461 mean, were generated using the FDA framework (**Figure 3A**). Let the muscle activity in
462 skeletal muscle m of subject i be the function value $x_{i,m}(t)$, and let the population mean
463 be $\mu_m(t)$ and the population variance be $\sigma_m(t) = \int_{i=1}^{n} \left(x_{i,m}(t) - \mu_m(t)\right)^2$. Functional
464 observations can be expressed using several orthogonal basis functions: $b_{m,q}(t); q =$
465 $1,2,\dots,Q$.

$$\mu_m(t) = \sum_{q=1}^{Q} \bar{c}_{m,q} b_{m,q}(t) \#(12)$$

466 This simulation used a B-spline basis, which is a typical orthogonal basis function in
467 the FDA. The coefficients $\bar{c}_{m,q}$ were estimated based on a regression spline and the number
468 of basis functions were determined by cross-validation. To mimic the actual muscle activity
469 pattern, we derive the mean function $\mu_m(t)$, which we would like to estimate, from a real
470 dataset. The average of the preprocessed muscle activity of the seven young healthy subjects
471 was calculated, and this was set as $\mu_m(t)$.
472 For the coefficient vector $\bar{\mathbf{c}}_m = \left(\bar{c}_{m,1}, \bar{c}_{m,2}, \dots, \bar{c}_{m,Q}\right)$, a subject-specific coefficient
473 vector $\mathbf{c}_{i,m} = \left(c_{i,m,1}, c_{i,m,2}, \dots, c_{i,m,Q}\right)$ was generated by adding an independently generated
474 noise normal distribution, which is considered biological variability (**Figure 4A**). That is, for
475 $q = 1,2,\dots,Q$,

$$\mathrm{c}_{i,m,q} = \bar{c}_{m,q} + \epsilon \#(13)$$

$$\epsilon \sim N(0, \sigma)$$

476 In this simulation, we set $\sigma = 1$ assuming that the subject population follows a standard

477 normal distribution. Finally, the orthogonal basis functions were multiplied again to generate

478 the functional observations for each subject.

$$x_{i,m}(t) = \sum_{q=1}^{Q} c_{i,m,q} b_{m,q}(t) \#(14)$$

479 We compared the algorithms for estimating the population mean of the motor modules using

480 muscle activity $\mathbf{X}_i = \{x_{t,m}\}$ consisting of $x_{i,t,m}$ discretized by t in equation (14).

481 For another simulation of the effects of outliers, heterogeneous samples were mixed

482 (**Figure 6A**). The muscle activity patterns of these samples were generated using equation

483 (13). That is, assume a different population of no interest with population mean $\mu_m^{'}(t)$;

$$\mu_m^{'}(t) = \sum_{q=1}^{Q} \bar{c}^{'}_{m,q} b_{m,q}(t) \#(15)$$

484 The coefficients $\bar{c}^{'}_{m,q}$ were randomly selected from the coefficients $\bar{c}_{m,q}$ of the population

485 mean of interest in equation (9), to which was added a noise of normal distribution with large

486 standard deviation $\sigma^{'}\ (> \sigma)$(**Figure 6A**);

$$\bar{c}^{'}_{m,q} = \bar{c}_{m,q} + \epsilon^{'} \#(16)$$

$$\epsilon \sim N\left(0, \sigma^{'}\right)$$

487 In this simulation on the effect of outlier, two coefficients $r_1, r_2\ (\leq Q)$) were randomly

488 selected from $q = 1,2, \ldots, Q$ and equation (13) was applied and obtained outlier samples;

$$x_{i^{'},m}(t) = \sum_{q=1}^{Q} c^{'}_{i^{'},m,q} b_{m,q}(t) \#(17)$$

The percentage of outlier samples was simulated as 10% of the total sample size, N.

Evaluation of methods

In this section, we evaluate the performance of these methods using a simulation dataset. We defined the motor module derived on the basis of equation (12) as the "true" motor module. Let $\mathbf{W}^{[true]}$ and $\mathbf{H}^{[true]}$ denote the true motor module estimates obtained by using the NMF algorithm. Each matrix element is denoted as $w_{tk}^{\text{true}} \in \mathbf{W}^{[true]}$ and $h_{km}^{\text{true}} \in \mathbf{H}^{[true]}$. Here, the primitive signal is represented as $\mathbf{W}^{[true]}$ and the weight of each muscle, that is, the module, is denoted by $\mathbf{H}^{[true]}$.

The residual sum of squares (RSS) of the true motor module was evaluated to assess the performance of each method. In other words, for the RSS of $k-$th synergy, we evaluated the

$$\text{RSS}_k^{[W]} = \sum_{t=1}^{T} \left(w_{tk}^{[true]} - w_{tk}^{[estimated]} \right)^2 \#(18)$$

$$\text{RSS}_k^{[H]} = \sum_{m=1}^{M} \left(h_{km}^{[true]} - h_{km}^{[estimated]} \right)^2 \#(19)$$

where $w_{tk}^{[estimated]}$ and $h_{km}^{[\text{estimated}]}$ are estimated values for $\mathbf{W}^{[true]}$ and $\mathbf{H}^{[true]}$, respectively.

Note that we normalized the values $w_{tk}^{[estimated]}$ and $h_{km}^{[\text{estimated}]}$ for each method because the scale of the estimated values differs from method to method. Normalization was performed for each estimated motor module by each method; each element was divided by

the value of maximum value in the estimated primitive signal $\mathbf{W}'^{[estimated]}$ and synergy $\mathbf{H}'^{[estimated]}$, and multiplied by 100. This was also true for $\mathbf{W}'^{[true]}$ and $\mathbf{H}'^{[true]}$. After normalization, these estimates were used to calculate the RSS using equations (18) and (19).

References

1. Ting, L.H., Chiel, H.J., Trumbower, R.D., Allen, J.L., McKay, J.L., Hackney, M.E., and Kesar, T.M. (2015). Neuromechanical principles underlying movement modularity and their implications for rehabilitation. Neuron *86*, 38–54.

2. Bizzi, E., and Cheung, V.C.K. (2013). The neural origin of muscle synergies. Front. Comput. Neurosci. *7*, 51.

3. Bizzi, E., Cheung, V.C.K., d'Avella, A., Saltiel, P., and Tresch, M. (2008). Combining modules for movement. Brain Res. Rev. *57*, 125–133.

4. d'Avella, A., Saltiel, P., and Bizzi, E. (2003). Combinations of muscle synergies in the construction of a natural motor behavior. Nat. Neurosci. *6*, 300–308.

5. Ting, L.H., and McKay, J.L. (2007). Neuromechanics of muscle synergies for posture and movement. Curr. Opin. Neurobiol. *17*, 622–628.

6. Tresch, M.C., and Jarc, A. (2009). The case for and against muscle synergies. Curr. Opin. Neurobiol. *19*, 601–607.

7. Safavynia, S.A., Torres-Oviedo, G., and Ting, L.H. (2011). Muscle Synergies: Implications for Clinical Evaluation and Rehabilitation of Movement. Top. Spinal Cord Inj. Rehabil. *17*, 16–24.

8. Tresch, M.C., Saltiel, P., and Bizzi, E. (1999). The construction of movement by the spinal cord. Nat. Neurosci. *2*, 162–167.

9. Godlove, J., Gulati, T., Dichter, B., Chang, E., and Ganguly, K. (2016). Muscle synergies after stroke are correlated with perilesional high gamma. Ann Clin Transl Neurol *3*, 956–961.

10. Ting, L.H., and Macpherson, J.M. (2005). A limited set of muscle synergies for force control during a postural task. J. Neurophysiol. *93*, 609–613.

11. Shourijeh, M.S., Flaxman, T.E., and Benoit, D.L. (2016). An approach for improving repeatability and reliability of non-negative matrix factorization for muscle synergy analysis. J. Electromyogr. Kinesiol. *26*, 36–43.

12. Cheung, V.C.K., and Seki, K. (2021). Approaches to revealing the neural basis of muscle synergies: a review and a critique. J. Neurophysiol. *125*, 1580–1597.

13. Rabbi, M.F., Pizzolato, C., Lloyd, D.G., Carty, C.P., Devaprakash, D., and Diamond, L.E. (2020). Non-negative matrix factorisation is the most appropriate method for extraction of muscle synergies in walking and running. Sci. Rep. *10*, 8266.

14. Turpin, N.A., Uriac, S., and Dalleau, G. (2021). How to improve the muscle synergy analysis methodology? Eur. J. Appl. Physiol. *121*, 1009–1025.

558 15. Chiovetto, E., Salatiello, A., d'Avella, A., and Giese, M.A. (2022). Toward a
559 unifying framework for the modeling and identification of motor primitives. Front. Comput.
560 Neurosci. *16*, 926345.

561 16. Amundsen Huffmaster, S.L., Van Acker, G.M., 3rd, Luchies, C.W., and Cheney, P.D.
562 (2018). Muscle Synergies Obtained from Comprehensive Mapping of the Cortical Forelimb
563 Representation Using Stimulus Triggered Averaging of EMG Activity. J. Neurosci. *38*,
564 8759–8771.

565 17. Tan, C.K., Kadone, H., Miura, K., Abe, T., Koda, M., Yamazaki, M., Sankai, Y., and
566 Suzuki, K. (2019). Muscle Synergies During Repetitive Stoop Lifting With a
567 Bioelectrically-Controlled Lumbar Support Exoskeleton. Front. Hum. Neurosci. *13*, 142.

568 18. Valk, T.A., Mouton, L.J., Otten, E., and Bongers, R.M. (2019). Fixed muscle
569 synergies and their potential to improve the intuitive control of myoelectric assistive
570 technology for upper extremities. J. Neuroeng. Rehabil. *16*, 6.

571 19. Ma, Y., Shi, C., Xu, J., Ye, S., Zhou, H., and Zuo, G. (2021). A Novel Muscle
572 Synergy Extraction Method Used for Motor Function Evaluation of Stroke Patients: A Pilot
573 Study. Sensors *21*. 10.3390/s21113833.

574 20. Cheung, V.C.K., Cheung, B.M.F., Zhang, J.H., Chan, Z.Y.S., Ha, S.C.W., Chen,
575 C.-Y., and Cheung, R.T.H. (2020). Plasticity of muscle synergies through fractionation and
576 merging during development and training of human runners. Nat. Commun. *11*, 4356.

577 21. Mehryar, P., Shourijeh, M.S., Rezaeian, T., Iqbal, N., Messenger, N., and
578 Dehghani-Sanij, A.A. (2017). Changes in synergy of transtibial amputee during gait: A pilot
579 study. In 2017 IEEE EMBS International Conference on Biomedical & Health Informatics
580 (BHI), pp. 325–328.

581 22. Mehryar, P., Shourijeh, M.S., Rezaeian, T., Khandan, A.R., Messenger, N., O'Connor,
582 R., Farahmand, F., and Dehghani-Sanij, A. (2020). Differences in muscle synergies between
583 healthy subjects and transfemoral amputees during normal transient-state walking speed. Gait
584 Posture *76*, 98–103.

585 23. Ferraty, F., and Vieu, P. (2006). Nonparametric Functional Data Analysis (Springer
586 New York).

587 24. Ramsay, J.O., and Silverman, B.W. (2005). Functional Data Analysis (Springer New
588 York).

589 25. Backenroth, D., Shinohara, R.T., Schrack, J.A., and Goldsmith, J. (2020).
590 Nonnegative decomposition of functional count data. Biometrics *76*, 1273–1284.

591 26. Berry, M.W., Browne, M., Langville, A.N., Pauca, V.P., and Plemmons, R.J. (2007).
592 Algorithms and applications for approximate nonnegative matrix factorization. Comput. Stat.
593 Data Anal. *52*, 155–173.

594 27. Santello, M., and Soechting, J.F. (2000). Force synergies for multifingered grasping.
595 Exp. Brain Res. *133*, 457–467.

596 28. Guan, N., Tao, D., Luo, Z., and Yuan, B. (2012). NeNMF: An Optimal Gradient
597 Method for Nonnegative Matrix Factorization. IEEE Trans. Signal Process. *60*, 2882–2898.

598 29. Kaminski, T.R. (2007). The coupling between upper and lower extremity synergies
599 during whole body reaching. Gait Posture *26*, 256–262.

600 30. Chiovetto, E., Patanè, L., and Pozzo, T. (2012). Variant and invariant features
601 characterizing natural and reverse whole-body pointing movements. Exp. Brain Res. *218*,
602 419–431.

603 31. Smith, C., Gilleard, W., Hammond, J., and Brooks, L. (2006). The Application of an
604 Exploratory Factor Analysis to Investigate the Inter-Relationships amongst Joint Movement
605 During Performance of a Football Skill. J. Sports Sci. Med. *5*, 417–524.

606 32. Steinberg, F., and Bock, O. (2013). Influence of cognitive functions and behavioral
607 context on grasping kinematics. Exp. Brain Res. *225*, 387–397.

608 33. Jolliffe, I.T., and Cadima, J. (2016). Principal component analysis: a review and
609 recent developments. Philos. Trans. A Math. Phys. Eng. Sci. *374*, 20150202.

610 34. Yuan, K.-H., and Zhong, X. (2008). 8. Outliers, Leverage Observations, and
611 Influential Cases in Factor Analysis: Using Robust Procedures to Minimize Their Effect.
612 Sociol. Methodol. *38*, 329–368.

613 35. Berrendero, J.R., Justel, A., and Svarc, M. (2011). Principal components for
614 multivariate functional data. Comput. Stat. Data Anal. *55*, 2619–2634.

615 36. Jacques, J., and Preda, C. (2014). Functional data clustering: a survey. Adv. Data
616 Anal. Classif. *8*, 231–255.

617 37. Zhang, M., and Parnell, A. (2022). Review of Clustering Methods for Functional
618 Data. arXiv [stat.ME].

619 38. Goudriaan, M., Papageorgiou, E., Shuman, B.R., Steele, K.M., Dominici, N., Van
620 Campenhout, A., Ortibus, E., Molenaers, G., and Desloovere, K. (2021). Muscle synergy
621 structure and gait patterns in children with spastic cerebral palsy. Dev. Med. Child Neurol. *64*,
622 462–468.

623 39. Kubota, K., Hanawa, H., Yokoyama, M., Kita, S., Hirata, K., Fujino, T., Kokubun, T.,
624 Ishibashi, T., and Kanemura, N. (2021). Usefulness of Muscle Synergy Analysis in

625 Individuals With Knee Osteoarthritis During Gait. IEEE Trans. Neural Syst. Rehabil. Eng. *29*,
626 239–248.

627 40. Hashiguchi, Y., Goto, R., and Naka, T. (2023). Effects of orthoses on muscle activity
628 and synergy during gait. PLoS One *18*, e0281541.

629 41. Cabaraux, P., Agrawal, S.K., Cai, H., Calabro, R.S., Casali, C., Damm, L., Doss, S.,
630 Habas, C., Horn, A.K.E., Ilg, W., et al. (2022). Consensus Paper: Ataxic Gait. Cerebellum.
631 10.1007/s12311-022-01373-9.

632 42. Tan, C.K., Kadone, H., Watanabe, H., Marushima, A., Hada, Y., Yamazaki, M.,
633 Sankai, Y., Matsumura, A., and Suzuki, K. (2020). Differences in Muscle Synergy Symmetry
634 Between Subacute Post-stroke Patients With Bioelectrically-Controlled Exoskeleton Gait
635 Training and Conventional Gait Training. Front Bioeng Biotechnol *8*, 770.

636 43. Alnajjar, F., Zaier, R., Khalid, S., and Gochoo, M. (2020). Trends and Technologies
637 in Rehabilitation of Foot Drop: A Systematic Review. Expert Rev. Med. Devices *18*, 31–46.

638 44. Lim, J., Lim, T., Lee, J., Sim, J., Chang, H., Yoon, B., and Jung, H. (2021).
639 Patient-specific functional electrical stimulation strategy based on muscle synergy and
640 walking posture analysis for gait rehabilitation of stroke patients. J. Int. Med. Res. *49*,
641 3000605211016782.

642 45. Michaud, F., Shourijeh, M.S., Fregly, B.J., and Cuadrado, J. (2020). Do Muscle
643 Synergies Improve Optimization Prediction of Muscle Activations During Gait? Front.
644 Comput. Neurosci. *14*, 54.

645 46. Mehrabi, N., Schwartz, M.H., and Steele, K.M. (2019). Can altered muscle synergies
646 control unimpaired gait? J. Biomech. *90*, 84–91.

647 47. Zhu, F., Kern, M., Fowkes, E., Afzal, T., Contreras-Vidal, J.-L., Francisco, G.E., and
648 Chang, S.-H. (2021). Effects of an exoskeleton-assisted gait training on post-stroke
649 lower-limb muscle coordination. J. Neural Eng. *18*. 10.1088/1741-2552/abf0d5.

650 48. Afzal, T., Zhu, F., Tseng, S.-C., Lincoln, J.A., Francisco, G.E., Su, H., and Chang,
651 S.-H. (2022). Evaluation of Muscle Synergy During Exoskeleton-Assisted Walking in
652 Persons With Multiple Sclerosis. IEEE Trans. Biomed. Eng. *69*, 3265–3274.

653 49. Zwaan, E., Becher, J.G., and Harlaar, J. (2011). Synergy of EMG patterns in gait as
654 an objective measure of muscle selectivity in children with spastic cerebral palsy. Gait
655 Posture *35*, 111–115.

656 50. VAN Criekinge, T., Saeys, W., Hallemans, A., Herssens, N., Lafosse, C., VAN Laere,
657 K., Dereymaeker, L., VAN Tichelt, E., DE Hertogh, W., and Truijen, S. (2020). SWEAT2
658 study: effectiveness of trunk training on muscle activity after stroke. A randomized controlled
659 trial. Eur. J. Phys. Rehabil. Med. *57*, 485–494.

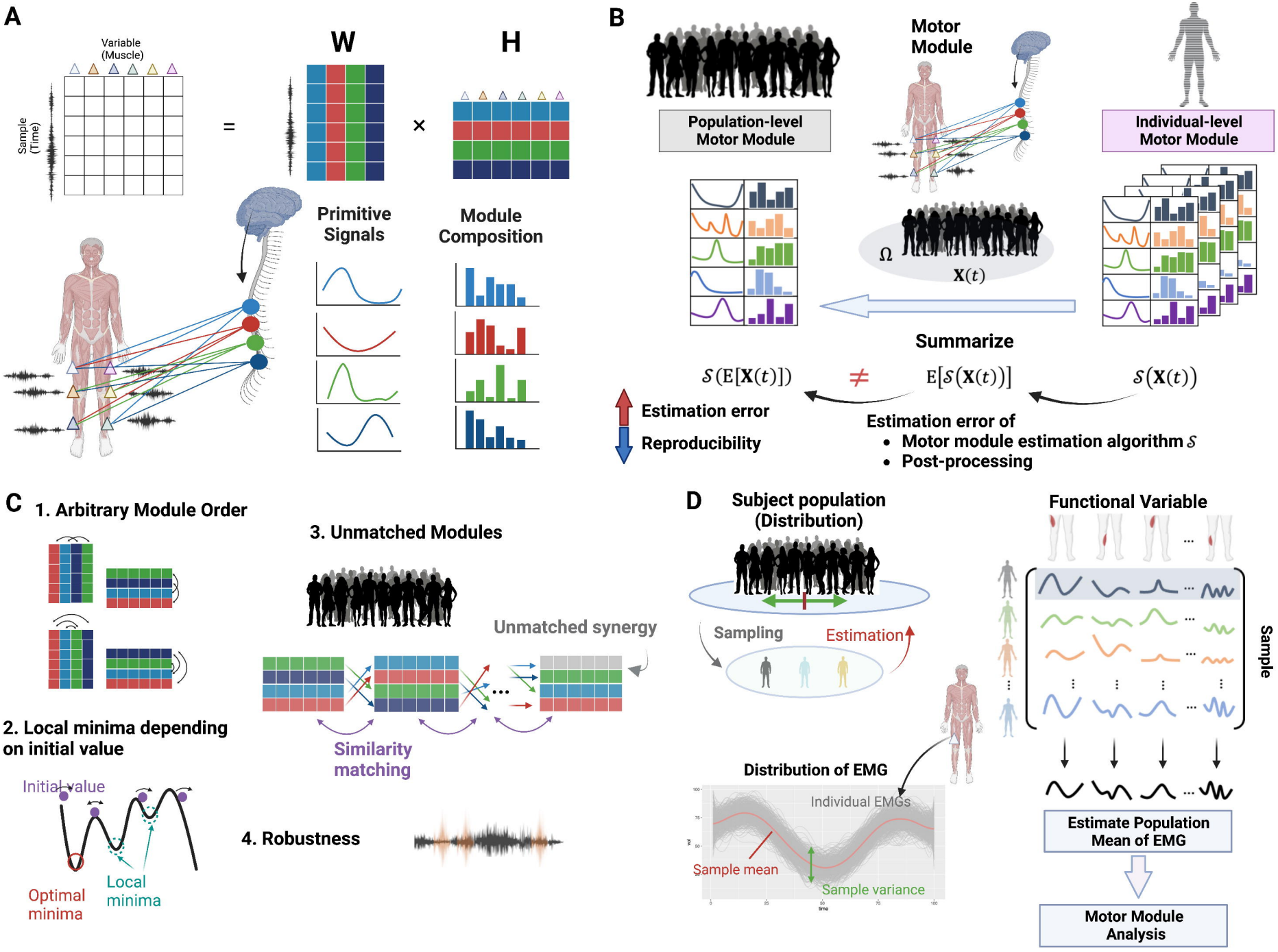
660 51. Acuña, S.A., Tyler, M.E., and Thelen, D.G. (2022). Individuals with Chronic
661 Mild-to-Moderate Traumatic Brain Injury Exhibit Decreased Neuromuscular Complexity
662 During Gait. Neurorehabil. Neural Repair *36*, 317–327.

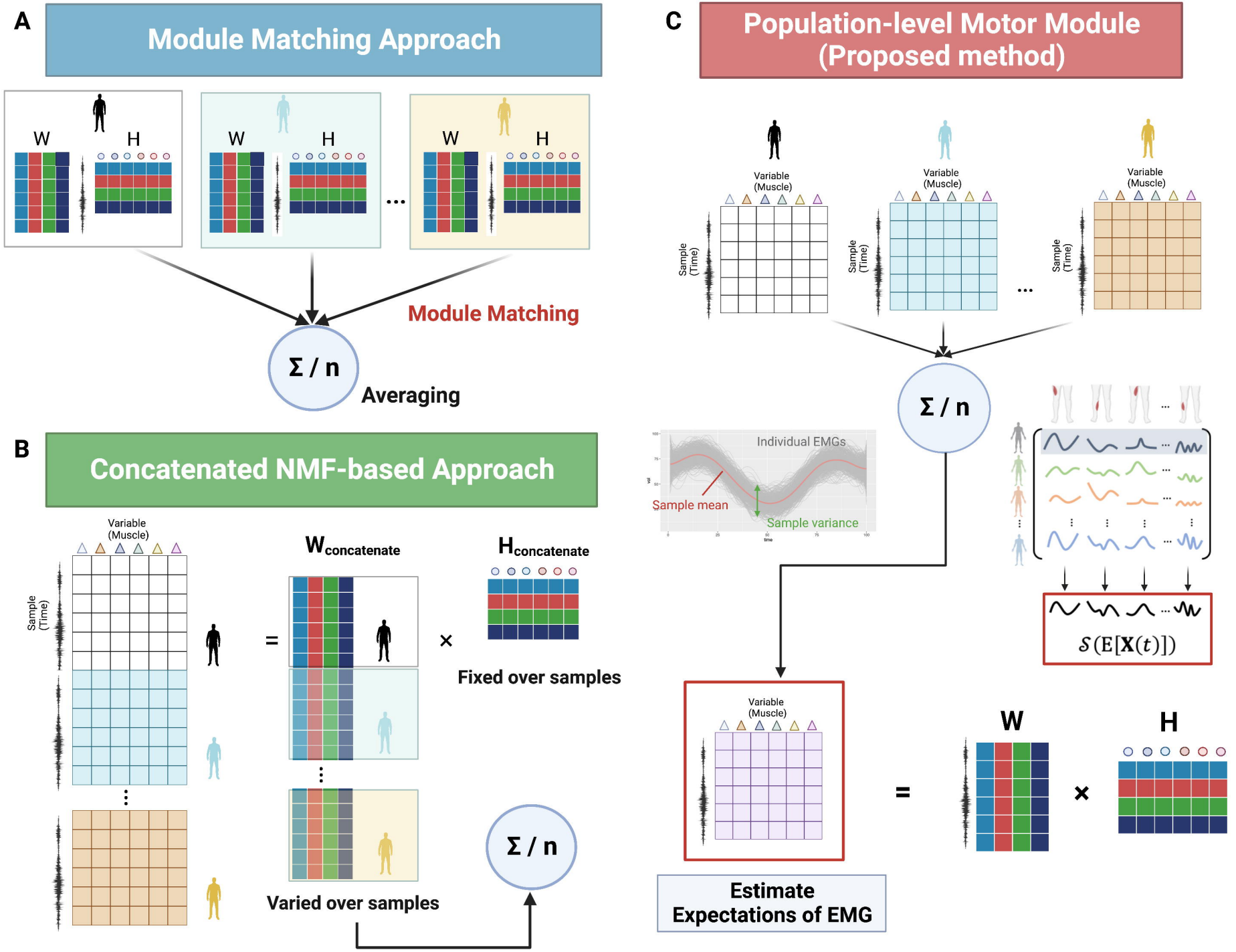
663 52. Hubert, M., Rousseeuw, P., and Verdonck, T. (2009). Robust PCA for skewed data
664 and its outlier map. Comput. Stat. Data Anal. *53*, 2264–2274.

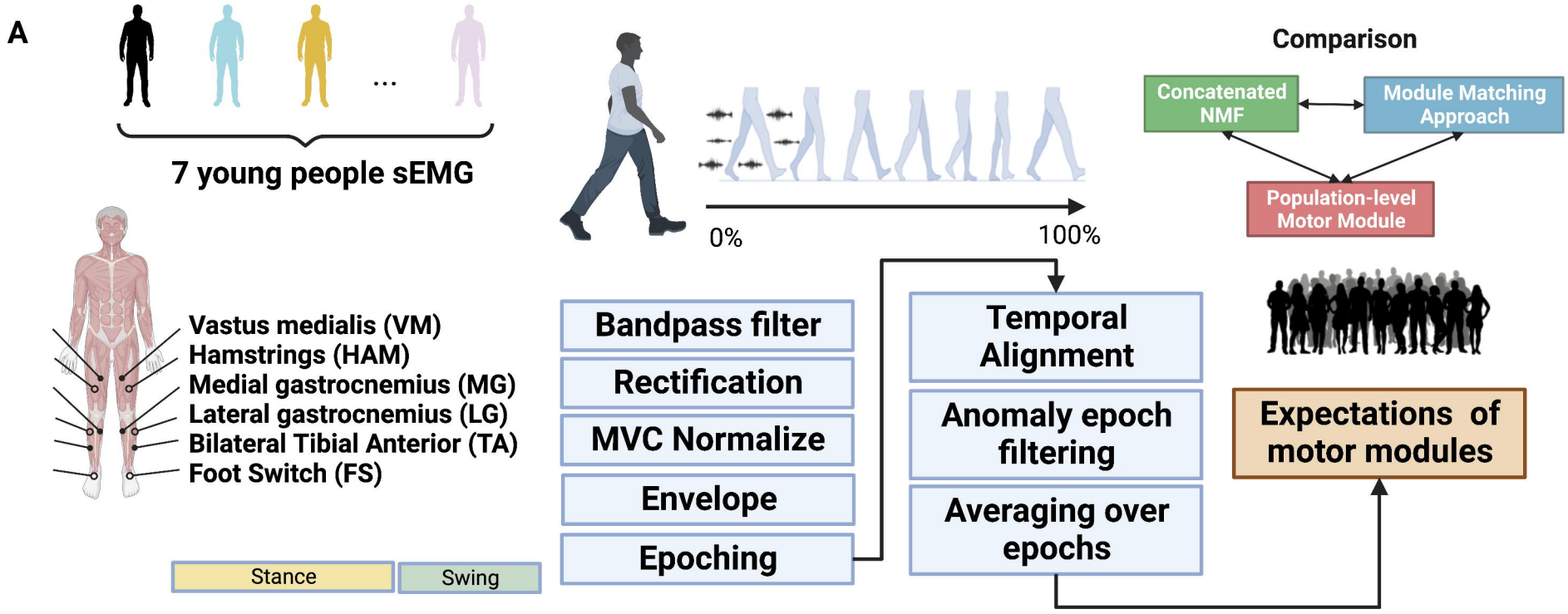
665 53. Engelen, S., Hubert, M., and Vanden Branden, K. (2005). A Comparison of Three
666 Procedures for Robust PCA in High Dimensions. AJS *34*, 117–126.

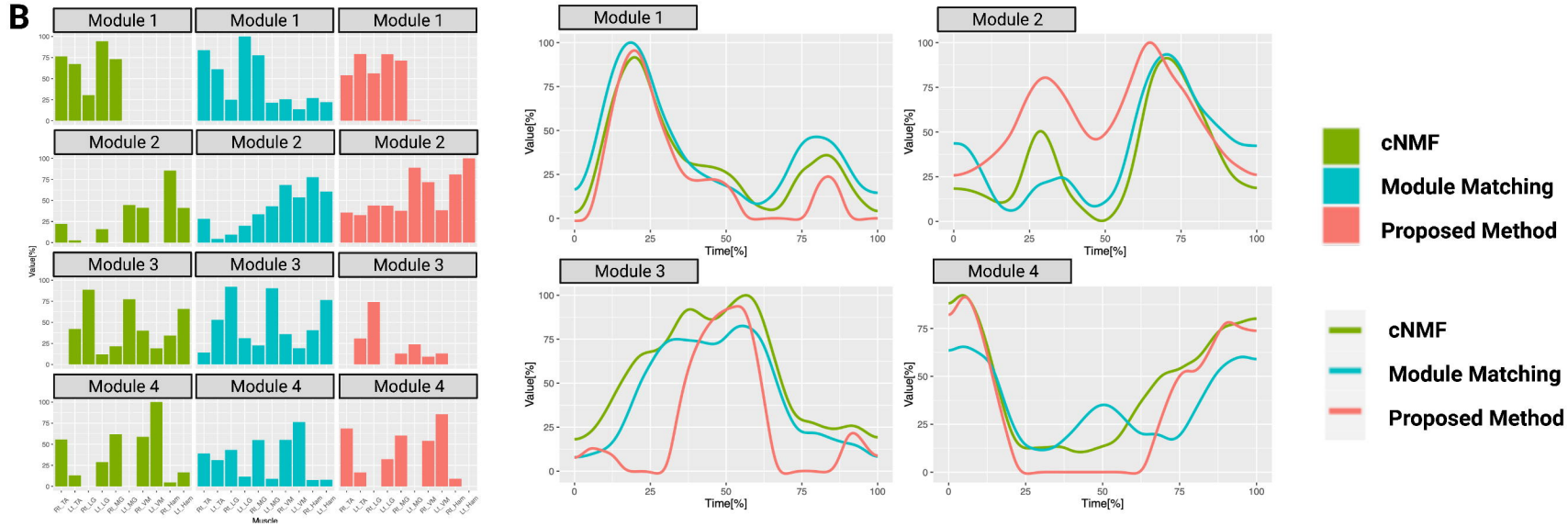
667 54. Hubert, M., Rousseeuw, P.J., and Vanden Branden, K. (2005). ROBPCA: A New
668 Approach to Robust Principal Component Analysis. Technometrics *47*, 64–79.

669 55. Reynkens, T. (2018). Robust Sparse PCA using the ROSPCA Algorithm [R package
670 rospca version 1.0.4].

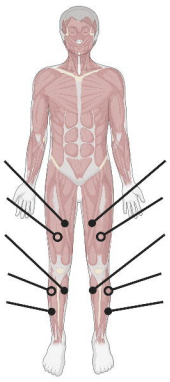

A
Variable (Muscle)
Sample (Time)
W
H
Primitive Signals
Module Composition
B
Population-level Motor Module
Motor Module
Individual-level Motor Module
Ω
X(t)
Summarize
S(E[X(t)])
≠
E[S(X(t))]
S(X(t))
Estimation error
Reproducibility
Estimation error of
Motor module estimation algorithm S
Post-processing
C
1. Arbitrary Module Order
2. Local minima depending on initial value
Initial value
Optimal minima
Local minima
3. Unmatched Modules
Unmatched synergy
Similarity matching
4. Robustness
D
Subject population (Distribution)
Sampling
Estimation
Distribution of EMG
Individual EMGs
Sample mean
Sample variance
Functional Variable
Sample
Estimate Population Mean of EMG
Motor Module Analysis


A
Module Matching Approach
W
H
Module Matching
Σ / n
Averaging
B
Concatenated NMF-based Approach
Variable (Muscle)
Sample (Time)
W concatenate
H concatenate
Fixed over samples
Varied over samples
Σ / n
C
Population-level Motor Module (Proposed method)
Variable (Muscle)
Sample (Time)
Σ / n
Individual EMGs
Sample mean
Sample variance
$\mathcal{S}(\mathrm{E}[\mathbf{X}(t)])$
Variable (Muscle)
W
H
Estimate Expectations of EMG


A
7 young people sEMG
Vastus medialis (VM)
Hamstrings (HAM)
Medial gastrocnemius (MG)
Lateral gastrocnemius (LG)
Bilateral Tibial Anterior (TA)
Foot Switch (FS)
0%
100%
Stance
Swing
Bandpass filter
Rectification
MVC Normalize
Envelope
Epoching
Temporal Alignment
Anomaly epoch filtering
Averaging over epochs
Comparison
Concatenated NMF
Module Matching Approach
Population-level Motor Module
Expectations of motor modules

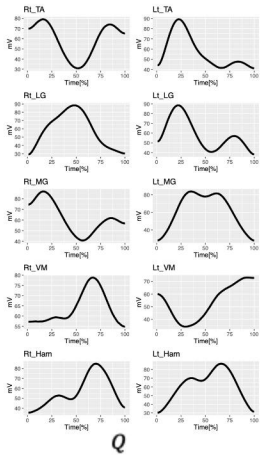

B
Module 1
Module 2
Module 3
Module 4
Value[%]
Time[%]
cNMF
Module Matching
Proposed Method

A

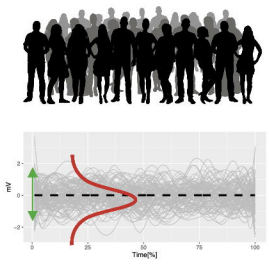


Vastus medialis (VM)
Hamstrings (HAM)
Medial gastrocnemius (MG)
Lateral gastrocnemius (LG)
Bilateral Tibial Anterior (TA)

True Mean of Muscle Activity



$$\mu_m(t) = \sum_{q=1}^{Q} \bar{c}_{m,q} b_{m,q}(t)$$

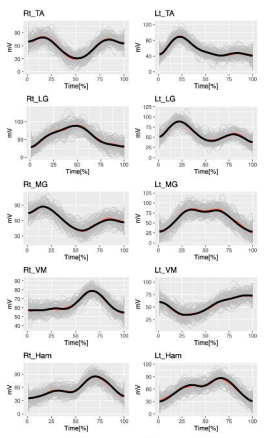


$$c_{i,m,q} = \bar{c}_{m,q} + \epsilon$$

$$\epsilon \sim N(0, \sigma)$$

Functional Gaussian noise

Random Muscle Activity

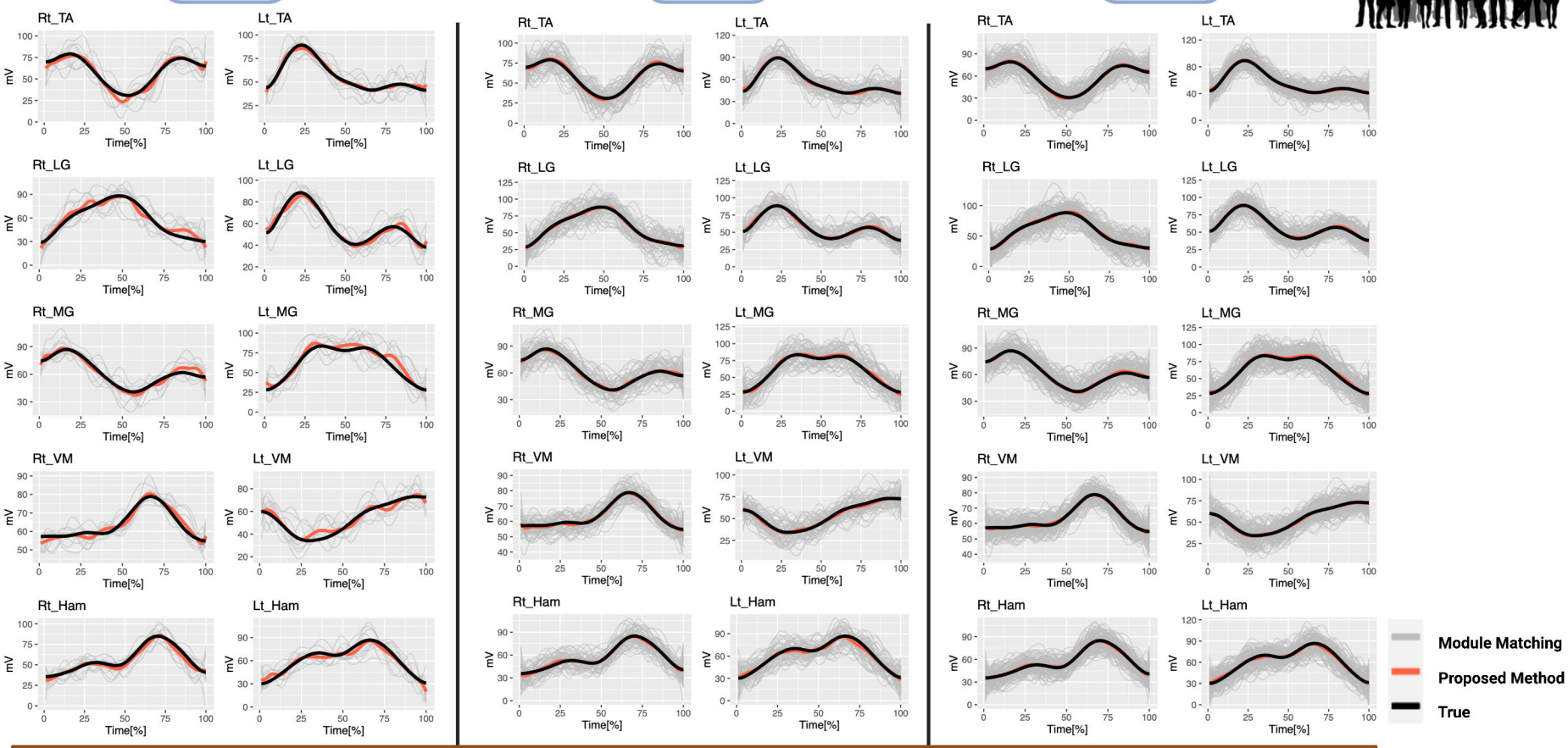


$$x_{i,m}(t) = \sum_{q=1}^{Q} c_{i,m,q} b_{m,q}(t)$$

Muscle activity

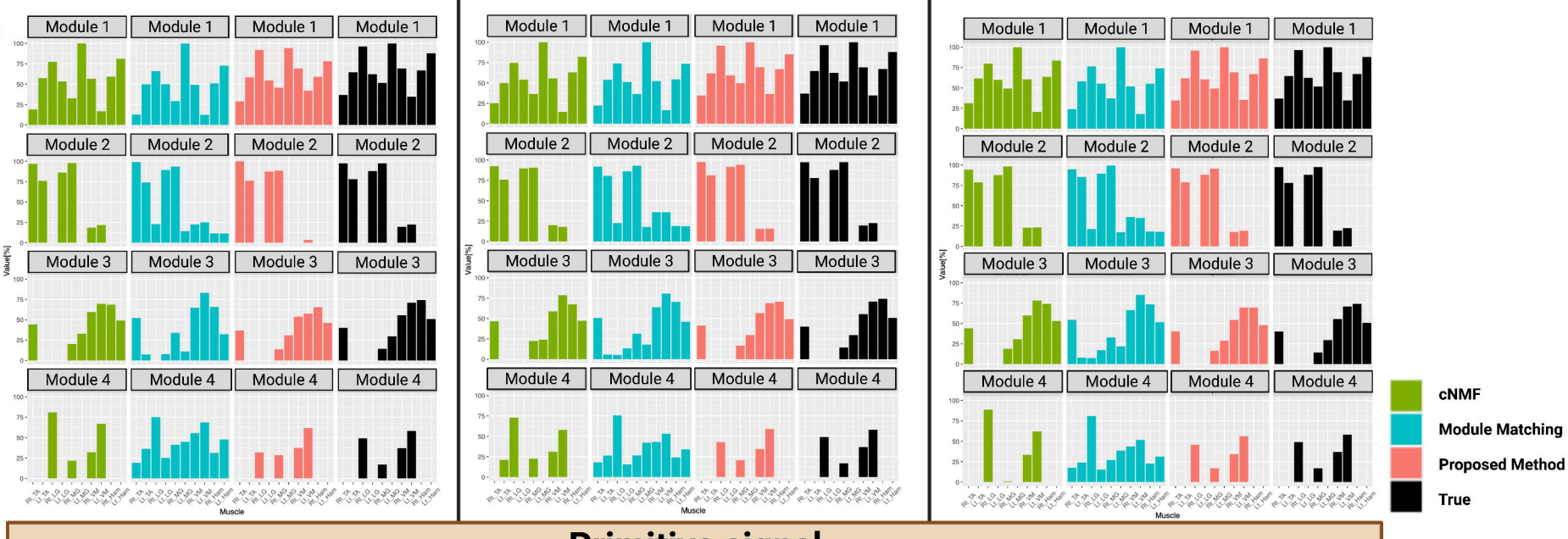
B

n=10 | n=50 | n=100

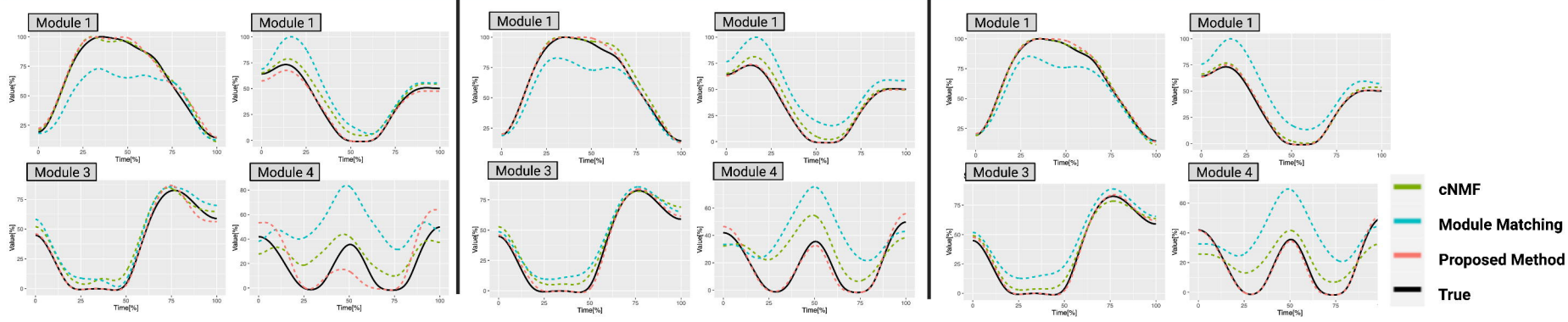


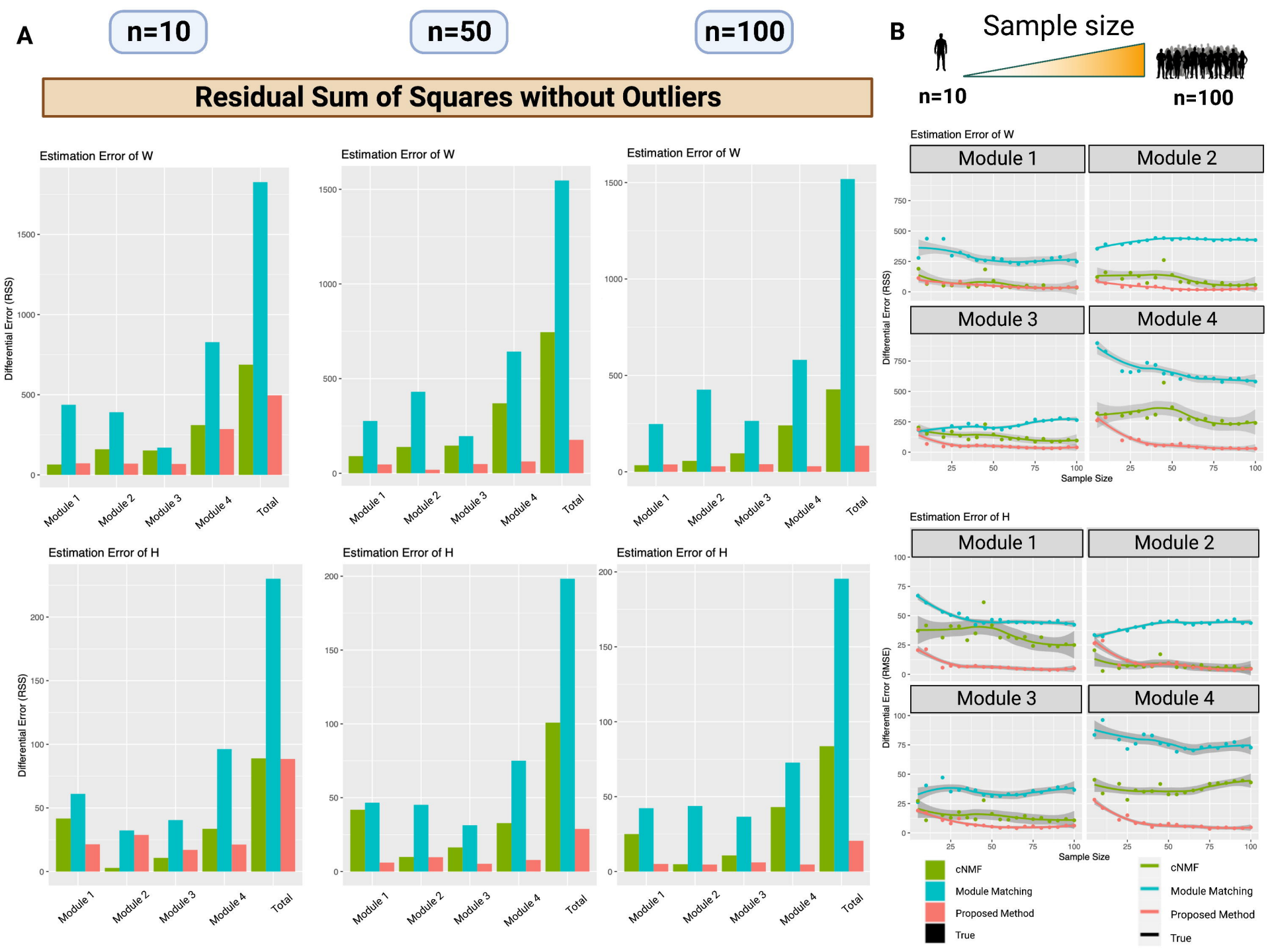
Module Composition

C



Primitive signal




A
n=10
n=50
n=100
B
Sample size
n=10
n=100
Residual Sum of Squares without Outliers
Estimation Error of W
Estimation Error of H
Differential Error (RSS)
Differential Error (RMSE)
Module 1
Module 2
Module 3
Module 4
Total
Sample Size
cNMF
Module Matching
Proposed Method
True

A

True Mean of Muscle Activity

Random Muscle Activity + Anomaly

$$c_{i,m,q} = \bar{c}_{m,q} + \epsilon$$

$$\epsilon \sim N(0, \sigma)$$

$$\bar{c}'_{m,q} = \bar{c}_{m,q} + \epsilon'$$

$$\epsilon \sim N(0, \sigma')$$

$$\mu_m(t) = \sum_{q=1}^{Q} \bar{c}_{m,q} b_{m,q}(t)$$

Functional Gaussian noise

Anomalous noise

$$x_{i,m}(t) = \sum_{q=1}^{Q} c_{i,m,q} b_{m,q}(t)$$

B

Muscle activity

n=10 n=50 n=100

Rt_TA Lt_TA Rt_LG Lt_LG Rt_MG Lt_MG Rt_VM Lt_VM Rt_Ham Lt_Ham

mV Time[%]

Module Matching
Proposed Method
True

C

Module Composition

Module 1 Module 2 Module 3 Module 4

Value[%] Muscle

cNMF
Module Matching
Proposed Method
True

Primitive signal

Module 1 Module 3 Module 4 Time[%]

cNMF
Module Matching
Proposed Method
True