Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spike Train Matching and Waveform Tracking Disagree: A Graph-Based Motor Unit Agreement Framework

Shiyu Zhang
University of Central Florida
Yue Wen
University of Central Florida
Mohsen Rakhshan
mohsen.rakhshan@ucf.edu
University of Central Florida

Research Article

Keywords: Motor Unit, sEMG Decomposition, Rate of Agreement, Waveform Tracking, Spike Train Matching, Network Analysis, Graph Theory

Posted Date: July 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10163973/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. Mohsen Rakhshan (the senior author) is an Editorial Board Member at JNER.

Spike Train Matching and Waveform Tracking Disagree: A Graph-Based Motor Unit Agreement Framework

Shiyu Zhang[1], Yue Wen[2,3*], Mohsen Rakhshan[1,3*]

[1]Department of Electrical and Computer Engineering, College of Engineering and Computer Science, University of Central Florida, Orlando, Florida, 32816, United States of America.
[2]Department of Mechanical and Aerospace Engineering, College of Engineering and Computer Science, University of Central Florida, Orlando, Florida, 32816, United States of America.
[3]Disability, Aging and Technology (DAT) Cluster, Faculty Cluster Initiative (FCI), University of Central Florida, Orlando, Florida, 32816, United States of America.

*Corresponding author(s). E-mail(s): yue.wen@ucf.edu;
mohsen.rakhshan@ucf.edu;
Contributing authors: shiyu.zhang@ucf.edu;

1 Abstract

1.1 Background

Motor units (MU) are the fundamental building blocks of movement, and reliably identifying the same MU across recordings is essential for neuroscience, rehabilitation, and neural interface research. Two mainstream methods for this purpose, spike train matching and waveform tracking, are widely used and often assumed to be interchangeable, yet this assumption has never been thoroughly tested.

1.2 Methods

We first decomposed MU activities from two independent datasets, utilizing the high-density surface electromyography (HDsEMG) decomposition technique. Then, to build up the agreement for comparing these two tracking methods using the MUs' activities, we decomposed them within the same experimental trial. The comparison was performed in two settings: first, the agreement between the two methods was evaluated using the real decomposed data and the simulated data. Then, the agreement between the two methods was evaluated using the real decomposed data.

1.3 Results

Analyzing over 14,000 motor units across two independent datasets, this study demonstrates that the two methods are not interchangeable, that their agreement depends critically on threshold selection and dataset characteristics, and that no single threshold works across datasets.

1.4 Conclusions

These findings demonstrate that the two methods are not interchangeable, that a high waveform similarity score alone does not guarantee MU identity, and that applying a fixed universal threshold is inappropriate. Dataset-specific calibration through a pilot study is necessary before either method is applied, with direct implications for prosthetics, rehabilitation, and neuromuscular disease research. These findings challenge decades of standard practice and establish that dataset-specific calibration is necessary

2 Keywords

Motor Unit, sEMG Decomposition, Rate of Agreement, Waveform Tracking, Spike Train Matching, Network Analysis, Graph Theory

3 Introduction

In the human peripheral nervous system (PNS), a MU consists of an alpha motor neuron and all the muscle fibers it innervates, creating the smallest functional unit of muscle force production. The electrophysiological activity of the MU, including motor unit spike trains (MUST) and motor unit action potential (MUAP) waveforms, serves

as critical intermediate information that can be used to assess the health and structural integrity of an individual's PNS and muscle[1,2]. Consequently, reliable identification and longitudinal tracking of the same MU can play a crucial role for neuroscience, kinesiology, and rehabilitation, as it helps in understanding how characteristics of a specific MU change in response to medical interventions[3], disease[4], or physical training[5].

To identify the same MU over time, researchers record electromyography (EMG) from either the target muscles directly (intra-muscular EMG, or iEMG) or the overlying skin surface (surface EMG, or sEMG), decompose the recordings to extract MUSTs, and obtain MUAP waveforms via spike-triggered averaging (STA)[6]. Upon decomposition, three methods can be used to identify the same MU across different time scales. Spike train matching[7] computes the ratio of overlapping spikes between two MUSTs within the same trial, determining MU uniqueness by comparing this ratio against a threshold. Filter tracking[3,8–10] reuses the separation vectors (also known as MU filters) learned during decomposition of one trial to extract MUs in other trials, enabling cross-trial MU identification. Waveform tracking[3,9,11–14] applies STA to obtain MUAP waveforms from each decomposed trial and determines MU identity by comparing waveform similarity scores against a threshold.

These three methods operate on different time scales, which determines when each can be applied. Spike train matching is generally restricted to within a single trial[15–17] because it relies on exact spike timing, which is unlikely to be preserved across trials[8,18]; it is therefore primarily used to identify and remove duplicate MUs within a trial[15,16,19–21] or to validate sEMG decomposition against iEMG decomposition in two-source validation protocols[15,17]. Filter tracking[10,22], conversely, can only be applied across trials, since separation vectors within a single trial are derived to produce non-identical MUs by construction. Waveform tracking is the most flexible, applicable both within[23] and across[12] trials. Consequently, within a single trial, researchers may use spike train matching[15,17] or waveform tracking[23]; across trials, they may use filter tracking[8,20], waveform tracking[11,13,14], or both[3,9]. These approaches can also be combined, where spike train matching first removes within-trial duplicates before waveform- or filter-based tracking is applied across trials[20].

Given that spike train matching and waveform tracking can both be applied within the same trial, it is unclear how their outcomes relate to one another. Two fundamental challenges have prevented researchers from establishing this connection. First, each method uses a distinct metric: spike train matching uses the rate of agreement (RoA)[15,16] to quantify the ratio of overlapping spikes between two MUSTs, while waveform tracking uses the cross-correlation coefficient (XCC)[8] to quantify MUAP waveform similarity. The correspondence between RoA and XCC is unknown. Second, both methods rely on a threshold to determine whether two MUs are the same, yet there is no consensus on appropriate values. Reported thresholds vary widely, from 30%[15,16,19–21] to 50%[24] or even 90%[7,17] for RoA, and from 0.7[25,26] to 0.8[11,13,14] or 0.9[9] for XCC. Moreover, applying a uniform threshold across all MUs ignores well-established variability in discharge patterns across MUs[27] and across task conditions[28], suggesting that fixed thresholds may be fundamentally inadequate.

Beyond the challenges of metric correspondence and threshold selection, both methods share a deeper limitation: pairwise comparisons of MUs are evaluated in isolation, discarding the broader connectivity structure that emerges when all decomposed MUs are considered simultaneously. This connectivity information can be naturally represented as a graph, in which nodes correspond to individual MUs and edges encode pairwise similarity (measured by RoA or XCC). Applying thresholds to the edges yields a sparse graph whose properties can be analyzed using network neuroscience tools, enabling the extraction of higher-order structures such as MU communities[29]. Such representations have proven useful for studying common synaptic input[30], MU functional connectivity[31], and duplicate removal[32]. Critically, this graph framework provides a principled basis for comparing the outcomes of spike train matching and waveform tracking not only locally, at the level of individual pairs, but globally, across the full structure of MU relationships.

In this paper, we examine how spike train matching and waveform tracking correspond to one another within a single trial, using both simulated and experimentally decomposed MU data. Leveraging the graph representation described above, we evaluate agreement at three levels: edge-level agreement, which quantifies pairwise correspondence between RoA and XCC; node-level agreement, which quantifies consistency in assigning individual MUs to the same group; and group-level agreement, which quantifies the overlap between MU communities identified by each method. Agreement is evaluated systematically across a range of threshold combinations for both methods. We demonstrate two principal findings. First, RoA and XCC are nonlinearly correlated, with XCC variability decreasing as RoA increases, revealing that the two metrics do not vary equivalently and cannot be used interchangeably. Second, the agreement between methods is both threshold-dependent and dataset-dependent, with consistent patterns observed across all three graph levels, indicating that applying fixed universal thresholds is inappropriate and that dataset-specific calibration is necessary before deploying either method. Together, these findings provide the first systematic evidence linking spike train matching and waveform tracking and present a graph-based framework for evaluating MU identification methods that is broadly applicable to future studies.

4 Methods

4.1 Dataset

Two open-source datasets were used in this study: the Hyser dataset[33] and the CEMHSEY dataset[34].

4.1.1 The 1DoF-Hyser dataset

The 1DoF and MVC sub-datasets of Hyser were used for this study (referred to as 1DoF henceforth). The MVC sub-dataset was used solely to normalize the load cell voltage readings into Maximum Voluntary Contraction (MVC) values; only the force recordings from this sub-dataset were used. Forces from both sub-datasets were recorded using an S-type bidirectional load cell (SAS Load Cell, Huatran, China) at

a sampling rate of 100 Hz. The 1DoF sub-dataset includes 256 channels of HD-sEMG (Quattrocento, OT Bioelettronica, Italy) and force recordings, with HD-sEMG sampled at 2048 Hz. Data were acquired from 20 subjects across two days, with each finger tested individually. Each finger underwent three repetitive contractions. Subjects performed a single-digit triangular isometric contraction for each trial, ranging from +30% MVC (extension) to −30% MVC (flexion) over 25 seconds. HD-sEMG recordings were collected from the extensor and flexor muscles of the right forearm using four 8×8 surface electrode arrays, each with a 10 mm inter-electrode distance. Two arrays were placed side by side on the extensor side of the forearm, and two on the flexor side with the same layout. A subset of the dataset was selected for analysis, comprising recordings from subjects 1 to 10, collected on both days, using channels from electrode grids on Extensor Digitorum Communis (EDC, channels 65-128) and on Flexor Digitorum Superficialis (FDS, channels 193-256).

4.1.2 The GRASP-CEMHSEY dataset

The GRASP sub-dataset of CEMHSEY was used for this study. It contains recordings from 10 subjects performing an isometric gripping task over 11 days, spanning 7 gripping gestures, 3 contraction levels, and 2 trials per contraction level. Force was recorded using a hand grip dynamometer (G200, Biometrics Ltd, United Kingdom) at 200 Hz. HD-sEMG was recorded using the same device and sampling rate as the 1DoF dataset. Details of electrode placement are provided in Yang et al.[34]. A subset of the dataset was selected for analysis: recordings involving the cylindrical grasp (session 1) collected on days 1 and 2, using channels from the electrode grids on EDC (channels 1-64) and on FDS (channels 129-192) for all 10 subjects.

4.2 MU activity and characteristic acquisition

Two types of MU activity, MUST and MUAP waveform, were acquired using HD-sEMG decomposition[16] and spike-triggered averaging[6], respectively.

4.2.1 HD-sEMG decomposition and MUST acquisition

A fast-ICA-based MU decomposition algorithm[35], combined with the CKC separation vector update technique[16,36], was used to decompose HD-sEMG signals into multiple discrete MUSTs, where each MUST corresponds to a single MU.

Prior to decomposition, the HD-sEMG signals were zero-phase filtered using a 4th-order IIR Comb Notch Butterworth filter to remove 50 Hz power-line interference, followed by zero-phase filtering using a 4th-order Bandpass Butterworth filter with a bandwidth of 20-500 Hz[30,37]. Automatic channel evaluation was then applied to remove the 5% worst channels, with the discard criterion based on the median correlation coefficient across all channels[38].

The filtered signal was whitened and decomposed with the following parameters: extension factor 17; ZCA whitening $\varepsilon = 0$; 50 iterations per contraction window; minimum total spike count of 40; SIL threshold ≥ 0.5. The PNR threshold and the coefficient of variation of the inter-spike interval (CoVISI) threshold were not applied. The SIL threshold was set to the minimum value reported in the literature[24], with

SIL values considered critical for evaluating decomposition separability[39]. Subsequent inspection confirmed that 95% of decomposed MUs had SIL values distributed between 0.84 and 0.99 (1DoF: $0.85 \leq \text{SIL} \leq 0.98$; GRASP: $0.84 \leq \text{SIL} \leq 0.99$). Each combination of subject, day, gripping gesture (if applicable), finger (if applicable), contraction level, trial, and muscle was decomposed independently. Each contraction segment starts with the first instance exceeding 1% MVC and ends with the last instance exceeding 1% MVC. Duplicate removal, expert inspection, and manual editing were not applied after decomposition.

4.2.2 MUAP waveform acquisition

Monopolar MUAP waveforms were computed using the spike-triggered averaging (STA) algorithm[6] within a 50-ms rectangular window centered on each spike, applied across all electrodes[3]. Longitudinal single-differential (LSD) MUAP waveforms were then obtained by applying the LSD spatial filter to the monopolar MUAP waveforms, computed by subtracting each channel's signal from that of its distal neighbor[40].

4.2.3 MU characteristic calculation

Three MU characteristics were computed from the decomposed MUST or MUAP waveform to provide additional validation information. The inter-spike interval (ISI) was first derived from the MUST as $\text{ISI}_i = t_{i+1} - t_i$, where t_i is the time index of the i-th spike. ISIs shorter than 25 ms (40 pps) or longer than 167 ms (6 pps) were excluded from further calculation[15].

The three characteristics are defined as follows. The instantaneous discharge rate (IDR, in peaks per second)[41] was computed as $\text{IDR} = 1/\text{ISI}$, measuring how rapidly an MU discharges. The coefficient of variation of the inter-spike interval (CoV_{ISI})[15] was computed as $\text{CoV}_{\text{ISI}} = \sigma_{\text{ISI}}/\mu_{\text{ISI}}$, where σ_{ISI} and μ_{ISI} are the standard deviation and arithmetic mean of the ISI, respectively; for humans performing isometric contractions, CoV_{ISI} is typically no larger than 0.3[42]. The peak-to-peak voltage of the MUAP waveform was computed for every electrode in the array; the maximum value across all electrodes was taken as the representative peak-to-peak value for that MU, which can be used to infer MU depth[43], or MU territory[44].

4.3 Low-level MU pair relations acquisition

4.3.1 Spike train matching

The RoA was used to quantify the similarity between a pair of MUSTs[16,36]:

$$\text{RoA} = \frac{|\mathcal{C}|}{|\mathcal{A}| + |\mathcal{B}| - |\mathcal{C}|} \times 100\% \tag{1}$$

where $\mathcal{A}$ and $\mathcal{B}$ denote the sets of spike time indices for MU A and MU B, respectively; $\mathcal{C} = \mathcal{A} \cap \mathcal{B}$ is the set of coincident spikes; and $|\cdot|$ denotes the cardinality of the set. A tolerance of ± 0.5 ms (± 1 sample at 2048 Hz) was applied when computing spike coincidences.

Prior to matching, lagging correction was applied by shifting one spike train relative to the other by the lag that maximized their correlation, computed over a range of -25 ms to $+25$ ms without jitter[19,37,45]. Two MUSTs were considered identical if their RoA exceeded a specified threshold. Commonly used values in the literature are 30%[15,16,19–21], 50%[24], and 90%[7,17].

4.3.2 Waveform tracking

The XCC of the flattened MUAP waveform was used to quantify the similarity between a pair of MUAP waveforms[8]. The LSD MUAP waveform of each MU was first flattened into a 1D vector along the muscle fiber direction. XCC was then computed between each pair of flattened waveforms with lagging correction applied within a ± 100 ms window, and the maximum XCC after lag correction was taken as the MUAP similarity value for that pair. Two MUAP waveforms were considered identical if their XCC exceeded a specified threshold. Commonly used values are 0.7[25,26], 0.8[11,13,14], and 0.9[9].

4.4 Higher order MU relational structure acquisition

Pairwise comparisons using spike train matching or waveform tracking assess whether two MUs are identical by evaluating their spike train overlap or MUAP waveform similarity. These pairwise similarity values, obtained through exhaustive comparison of all MU pairs within a decomposed MU set, can be naturally represented as a weighted graph[29], where each node corresponds to an MU and each edge encodes the similarity between the connected pair (Fig. 1a). This graph-based representation captures the full relational structure of the MU set, enabling analysis not only at the level of individual pairs but also at the level of MU groups and communities.

Three graph representations were used to encode different aspects of this relational structure. The weighted graph encodes all pairwise similarity values as continuous numbers between 0 and 1. The binary graph encodes whether each pair is considered identical (1) or distinct (0), obtained by thresholding the weighted graph. The connected components of the binary graph encode how MUs cluster into groups, with each component representing a set of mutually connected MUs.

4.4.1 Decomposed MU set

The higher-order relational structure was computed within a decomposed MU set, defined as the collection of MUs obtained from decomposing the HD-sEMG recording for a specific experimental condition (subject, day, gripping gesture if applicable, finger if applicable, contraction level, trial, and muscle). After decomposition, 600 decomposed MU sets were obtained from the 1DoF dataset and 240 sets from the GRASP dataset.

4.4.2 Comparison matrix and weighted graph

After performing spike train matching or waveform tracking between all possible MU pairs within a decomposed MU set, the results were organized into a square symmetric

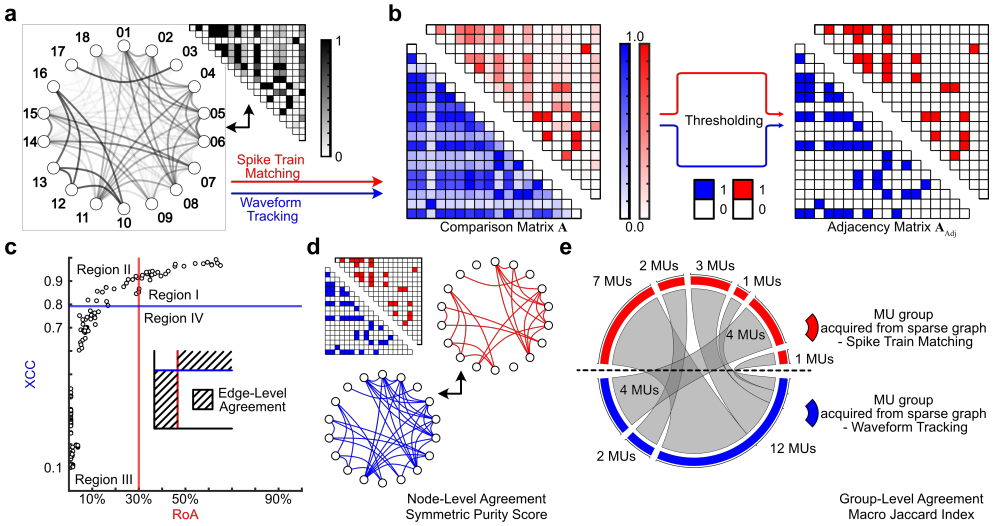


Fig. 1 Paradigm for graph-based data acquisition and agreement quantification using the real decomposed MUs. **a.** Pairwise comparisons within the decomposed MU set yield a symmetric comparison matrix, which corresponds directly to a weighted graph. Circles represent individual MUs (nodes); edges encode pairwise similarity (RoA or XCC). Red: spike train matching; Blue: waveform tracking. Self-comparisons (diagonal) are set to 0. Only the upper triangular portion of the matrix is shown. **b. left** RoA comparison matrix (red) and XCC comparison matrix (blue). Matrices are color-coded; darker colors indicate higher similarity values. **b. middle** Thresholding operation converting the comparison matrix into a binary adjacency matrix. Different threshold values yield different adjacency matrices. **b. right** Adjacency matrix. A colored block indicates a valid connection (identical MU pair); a white block indicates no connection (distinct MU pair). **c.** Jointly plotting RoA and XCC from pairwise comparisons enables examination of how spike train matching and waveform tracking remove edges differently. Red vertical line: RoA threshold. Blue horizontal line: XCC threshold. Four regions are defined by the two thresholds. Region 1: RoA $\geq$ $\text{RoA}_{\text{Threshold}}$ and XCC $\geq$ $\text{XCC}_{\text{Threshold}}$ (agreement: both methods identify the pair as identical). Region 2: RoA $<$ $\text{RoA}_{\text{Threshold}}$ and XCC $\geq$ $\text{XCC}_{\text{Threshold}}$ (disagreement: waveform tracking identifies as identical, spike train matching does not). Region 3: RoA $<$ $\text{RoA}_{\text{Threshold}}$ and XCC $<$ $\text{XCC}_{\text{Threshold}}$ (agreement: both methods identify the pair as distinct). Region 4: RoA $\geq$ $\text{RoA}_{\text{Threshold}}$ and XCC $<$ $\text{XCC}_{\text{Threshold}}$ (disagreement: spike train matching identifies as identical, waveform tracking does not). Edge-level agreement is defined as the fraction of pairwise comparisons in Regions 1 and 3. **d, e.** Each adjacency matrix corresponds to a binary graph. Cluster analysis is used to assess graph similarities at the node level (d, symmetric purity score) and group level (e, macro Jaccard index).

matrix, referred to as the comparison matrix (Fig. 1a). The number of rows and columns equals the number of MUs in the set, and each entry $\mathbf{A}_{ij}$ encodes the RoA (or XCC) between the i-th and j-th MU as a value between 0 and 1. The diagonal elements are set to 0, since self-comparisons are excluded. Each comparison matrix corresponds directly to a weighted graph [29], in which each row (or column) corresponds to a node and each off-diagonal entry corresponds to the weight of the edge connecting the corresponding pair of nodes.

4.4.3 Adjacency matrix and binary graph

A binary adjacency matrix was obtained by thresholding the comparison matrix: entries exceeding the specified threshold were set to 1 (indicating an identical MU pair), and all others were set to 0 (indicating a distinct pair) (Fig. 1b). The diagonal elements of the adjacency matrix remain 0. Each row (or column) of the adjacency matrix corresponds to a node in the binary graph, and each entry encodes the presence or absence of an edge, thereby encoding the graph's connectivity structure. Because this thresholding operation is equivalent to the decision step in spike train matching or waveform tracking, the adjacency matrix directly represents the output of either method at a given threshold.

4.4.4 Connected components and the MU group acquisition

MU groups were identified by extracting the connected components of the binary graph[29]. Two MUs connected by an edge in the binary graph are assigned to the same group, and the full set of connected MUs forms a group representing a set of mutually identified MUs. This procedure can be regarded as a form of clustering: both spike train matching and waveform tracking implicitly partition the decomposed MU set into groups, with the specific partition depending on the method and threshold applied.

4.5 Analysis framework

4.5.1 Correlation study between RoA and XCC using simulated data

To study how the RoA used in spike train matching and the XCC used in waveform tracking relate to one another, both metrics must be computed from MU pairs of known identity. Because both the 1DoF and GRASP datasets contain only sEMG recordings without concurrent iEMG ground truth, MU identity cannot be independently verified from the data. To overcome this, artificial MUSTs were generated from each real decomposed MUST by controlling RoA at nine levels (10% to 90% in 10% increments). The XCC between each artificial MU and its reference was then computed, with MU identity known by construction.

Artificial MU generation. The artificial MUST was generated by selecting a subset of spikes from the real decomposed MUST and adding spikes at time locations where the reference MU did not fire, in proportions determined by the target RoA. Because the total number of spikes in the artificial MUST affects the RoA calculation (see Section 4.3.1), two generation approaches were considered.

In the surrogate spike approach, the total number of spikes in the artificial MUST was constrained to equal that of the reference decomposed MUST. Spikes were drawn from the reference MUST to fill the shared portion, and the remaining spikes were placed at time locations where the reference MU did not fire.

In the subset spike approach, the artificial MUST consisted solely of a subset of the reference spikes, with no additional spikes added. This approach provides an

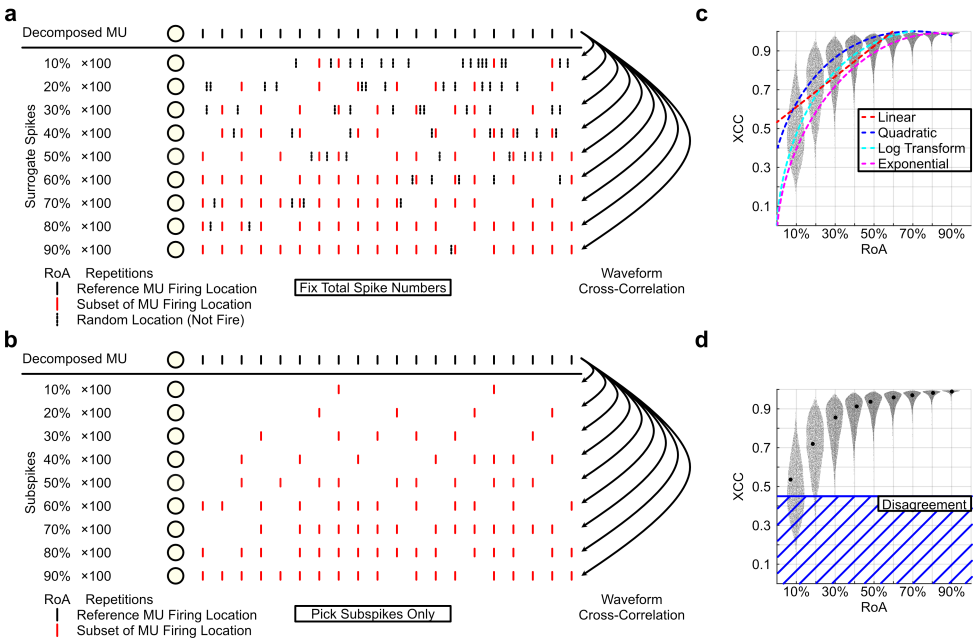


Fig. 2 Paradigm for using simulated data to study the RoA-XCC correlation and evaluate agreement between spike train matching and waveform tracking. **a, b** Artificial MU generation and MUAP waveform XCC calculation between real decomposed MUs and generated artificial MUs. RoA was controlled from 10% to 90% in 10% increments. Two approaches were considered: (a) surrogate spike approach, in which the artificial MU has the same total number of spikes as the reference decomposed MUST; and (b) subset spike approach, in which the artificial MU consists of a strict subset of the reference spikes. For every MU, artificial MU generation was repeated 100 times, and the median XCC across repetitions was used as the representative value. **c** Data fitting between RoA and XCC for each MU using four candidate models: linear mixed-effects (LME, red) model, LME model with a quadratic term (blue), log-transformed LME model (cyan), and a nonlinear mixed-effects (NLME) model with an exponential structure (magenta). Each dot represents the median XCC computed for a specific MU at a specific RoA level. **d** Agreement quantification between spike train matching and waveform tracking on the simulated data. Agreement at each RoA level was defined as the proportion of artificial MUs whose XCC exceeded the waveform-tracking threshold. All artificial MUs are true-positive matches by construction; data points below the threshold represent cases where waveform tracking disagrees with spike train matching.

upper bound on the XCC achievable at each RoA level, since all artificial spikes are guaranteed to be true spikes from the reference MU.

Artificial MUs were generated for every decomposed MU (1DoF: 10,429 MUs; GRASP: 4,296 MUs) at all nine RoA levels and for both approaches, with 100 repetitions per MU per level.

MUAP Waveform XCC between real and simulated MUs. LSD MUAP waveforms were computed for all MUs in the three sets: the original decomposed MUs, the surrogate spike approach artificial MUs, and the subset spike approach artificial MUs. XCC was then calculated between the original decomposed MUs and

each artificial set (Fig. 2a, b). The median XCC across 100 repetitions was used as the representative value for each MU at each RoA level.

Data fitting and model selection. To characterize the relationship between RoA and XCC, four candidate models incorporating MU-specific random effects were fitted to the data. Let i index the MU and j index the RoA level within that MU.

The LME model is (namely, linear):

$$\mathrm{XCC}_{ij} = (\beta_1 + b_{1i})\,\mathrm{RoA}_{ij} + (\beta_0 + b_{0i}) + \epsilon_{ij} \tag{2}$$

The LME model with a quadratic term is (namely, quadratic):

$$\mathrm{XCC}_{ij} = (\beta_2 + b_{2i})\,\mathrm{RoA}_{ij}^2 + (\beta_1 + b_{1i})\,\mathrm{RoA}_{ij} + (\beta_0 + b_{0i}) + \epsilon_{ij} \tag{3}$$

The log-transformed LME model is (namely, log-transformed):

$$\mathrm{XCC}_{ij} = (\beta_1 + b_{1i})\,\ln(\mathrm{RoA}_{ij}) + (\beta_0 + b_{0i}) + \epsilon_{ij} \tag{4}$$

The NLME model with an exponential structure is (namely, exponential):

$$\mathrm{XCC}_{ij} = 1 - \exp[(\beta + b_i)\,\mathrm{RoA}_{ij}] + \epsilon_{ij} \tag{5}$$

In all models, β (with subscripts where applicable) denotes population-level fixed-effect coefficients, b_i (with subscripts where applicable) denotes MU-specific random effects, and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ denotes residual error. For the LME models, random effects on both intercept and slope terms are assumed jointly normal with mean zero and variance-covariance matrix $\mathbf{D}$. For the NLME model, the scalar random effect is assumed $b_i \sim \mathcal{N}(0, \psi)$. Model selection was performed using the Bayesian Information Criterion (BIC), with lower values indicating a better trade-off between goodness-of-fit and model complexity.

4.5.2 Quantifying agreement on simulated data

Because the artificial MUSTs were generated from real decomposed MUSTs at controlled RoA levels, every artificial MU is a true-positive match with its reference by construction. Agreement between spike train matching and waveform tracking was therefore quantified as the proportion of artificial MU pairs at each RoA level for which XCC exceeded the waveform-tracking threshold (Fig. 2d). Artificial MUs whose XCC falls below the threshold are counted as disagreements: spike train matching correctly identifies them as identical to the reference, whereas waveform tracking does not.

4.5.3 RoA and XCC distribution analysis using decomposed MUs

For each pair of decomposed MUs within a MU set, RoA was computed via spike train matching and XCC via waveform tracking. The joint distribution of RoA and XCC was constructed by plotting RoA on the x-axis and XCC on the y-axis for all pairwise comparisons (Fig. 1c). All pairwise comparisons were aggregated into a single

overlaid plot (Supplementary Fig. **??**), and the marginal distributions of RoA and XCC were obtained by projecting this joint distribution onto the respective axes.

Evaluation of dataset-dependent pattern To examine whether the marginal or the joint distribution exhibits dataset-dependent patterns, we conducted a series of evaluations. We first examine whether the marginal distributions exhibit dataset-dependent patterns by comparing the RoA (or XCC) marginal distributions between the two datasets. After determining whether the marginal distributions exhibit dataset-dependent patterns, we then examined whether the correlation between RoA and XCC also differs across datasets. To this end, Pearson correlation coefficients (r_{1DoF} and r_{GRASP}) between RoA and XCC were first computed for each dataset. Then we examine whether this correlation shows a significant difference between the two datasets. Specifically, the correlation coefficients were first transformed using Fisher's z-transformation ($z = \frac{1}{2}\ln(\frac{1+r}{1-r})$), and the difference between the transformed values was evaluated using a z-statistic with standard error $SE = \sqrt{\frac{1}{n_{\mathrm{1DoF}}-3} + \frac{1}{n_{\mathrm{GRASP}}-3}}$, where n_{1DoF} and n_{GRASP} denote the number of paired observations used to compute the correlation in each dataset. Finally, the corresponding p-value was obtained from the standard normal distribution using a two-tailed test.

Evaluation of the model size of the marginal distribution of RoA or XCC To assess modal size, we first use the Hartigan dip test[46] to test whether each marginal distribution violates unimodality. When unimodality was rejected, GMMs with 1 to 6 components were fitted to the data to estimate the number of modes. The optimal number of components was selected using BIC. During fitting, a regularization value of 10^{-6} and a maximum of 1000 iterations were used, with each model initialized 20 times. The modality of the distributions is directly relevant to whether thresholds for spike train matching or waveform tracking can be determined analytically from distributional properties[7,17].

4.5.4 Quantifying edge-level agreement using decomposed MUs

Jointly plotting RoA and XCC for all pairwise comparisons defines a two-dimensional plane that can be divided into four regions by the spike train matching threshold (a vertical line at the RoA threshold) and the waveform-tracking threshold (a horizontal line at the XCC threshold) (Fig. 1c):

Region 1: $\mathrm{RoA} \geq \mathrm{RoA}_{\mathrm{Threshold}}$ and $\mathrm{XCC} \geq \mathrm{XCC}_{\mathrm{Threshold}}$. Both methods identify the pair as identical (agreement).

Region 2: $\mathrm{RoA} < \mathrm{RoA}_{\mathrm{Threshold}}$ and $\mathrm{XCC} \geq \mathrm{XCC}_{\mathrm{Threshold}}$. Waveform tracking identifies the pair as identical; spike train matching does not (disagreement).

Region 3: $\mathrm{RoA} < \mathrm{RoA}_{\mathrm{Threshold}}$ and $\mathrm{XCC} < \mathrm{XCC}_{\mathrm{Threshold}}$. Both methods identify the pair as distinct (agreement).

Region 4: $\mathrm{RoA} \geq \mathrm{RoA}_{\mathrm{Threshold}}$ and $\mathrm{XCC} < \mathrm{XCC}_{\mathrm{Threshold}}$. Spike train matching identifies the pair as identical; waveform tracking does not (disagreement).

Edge-level agreement was defined as the proportion of pairwise comparisons falling in Regions 1 or 3:

$$\text{Agreement}_{\text{edge}} = \frac{\#\text{pairs in Regions 1 and 3}}{\#\text{total pairs}} \tag{6}$$

To examine how threshold selection affects edge-level agreement, both the spike train matching threshold and the waveform-tracking threshold were searched from 0.01 to 0.99 in increments of 0.01, and edge-level agreement was computed at every threshold combination.

4.5.5 Quantifying node-level agreement

Edge-level agreement quantifies pairwise consistency but does not directly reflect how many MUs are assigned to the same group by both methods. Node-level agreement was therefore defined as the average proportion of MUs clustered into the same group by both methods, and was quantified using the symmetric purity score[47].

Given a decomposed MU set of N MUs, spike train matching yields a partition into classes $\mathscr{C} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_r\}$ and waveform tracking yields a partition into clusters $\mathscr{K} = \{\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_s\}$, where r and s are the total number of groups in each partition. The symmetric purity score $\rho_D(\mathscr{C}, \mathscr{K})$ is defined as the product of the purity score $\mathcal{P}(\mathscr{C}, \mathscr{K})$ and the inverse purity score $\mathcal{P}(\mathscr{K}, \mathscr{C})$:

$$\rho_D(\mathscr{C}, \mathscr{K}) = \mathcal{P}(\mathscr{C}, \mathscr{K}) \times \mathcal{P}(\mathscr{K}, \mathscr{C}) \tag{7}$$

where

$$\mathcal{P}(\mathscr{C}, \mathscr{K}) = \frac{1}{N} \sum_i \max_j |\mathcal{C}_i \cap \mathcal{K}_j| \tag{8}$$

$$\mathcal{P}(\mathscr{K}, \mathscr{C}) = \frac{1}{N} \sum_j \max_i |\mathcal{C}_i \cap \mathcal{K}_j| \tag{9}$$

The purity score measures how well the clusters yielded by waveform tracking agree with the classes from spike train matching; the inverse purity score measures the converse. Both spike train matching and waveform tracking thresholds were searched from 0.05 to 0.95 in increments of 0.05, and node-level agreement was computed at each threshold combination. The median symmetric purity score across all decomposed MU sets was used as the representative node-level agreement value at each threshold combination.

4.5.6 Quantifying group-level agreement

Node-level agreement, as quantified by the symmetric purity score, includes duplicate MUs as part of the MU groups. Researchers may therefore question whether this measure fully reflects agreement when only distinct MUs are considered, since a decomposition with many duplicates could still yield high purity. To address this, group-level agreement was defined as the average overlap between unique MU groups identified by the two methods, and was quantified using the macro Jaccard index[48–50].

The macro Jaccard index computes the Jaccard index for each pair of matched groups, then takes an equally weighted average across all groups. Prior to computation, MU groups identified by spike train matching and those identified by waveform tracking were aligned by maximizing the number of overlapping MUs between corresponding group pairs. The Jaccard index for a matched pair of groups A and B is:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{10}$$

The macro Jaccard index is then:

$$J_{\text{Macro}} = \frac{1}{C} \sum_{i} J_i, \quad C = \max(r, s) \tag{11}$$

where C is the maximum number of groups across the two methods, ensuring that unmatched groups contribute a Jaccard index of 0 to the average. Both thresholds were searched from 0.05 to 0.95 in increments of 0.05, and the median macro Jaccard index across all decomposed MU sets was used as the representative group-level agreement value at each threshold combination.

4.6 Statistical analysis

All statistical analyses were performed in MATLAB (MathWorks, Natick, MA; Version R2024b). The significance level was set at $\alpha = 0.05$. All data were initially tested for normality using the Shapiro-Wilk test[51]. For normally distributed data, results are reported as mean ± standard deviation. For non-normally distributed data, results are reported as median ± mean absolute deviation (MAD).

For pairwise comparisons, a paired t-test was used for paired normally distributed data and a Wilcoxon signed-rank test for paired non-normal data. For unpaired comparisons, an unpaired t-test was used for normally distributed data and a Wilcoxon rank-sum test for non-normal data. When comparing a single group against a specified value, a one-sample t-test or one-sample Wilcoxon test was applied depending on the normality of the data. For comparisons among more than two groups, repeated-measures ANOVA with Tukey's HSD post hoc test was used for normally distributed data, and the Kruskal-Wallis test with Dunn's post hoc test for non-normal data. Where multiple comparisons were performed, p-values were corrected using the Bonferroni correction.

For data with non-unimodal distributions, the Hartigan dip test[46] was used to confirm non-unimodality. Following confirmation, the optimal number of modes was estimated by fitting GMM with increasing numbers of components and selecting the component count that minimized the BIC.

5 Results

5.1 Relation examination and agreement calculation between RoA and XCC using simulated data

Understanding how spike train matching and waveform tracking correspond requires first establishing how their respective metrics, the spike train RoA and the MUAP waveform XCC, relate to one another. A strong linear correlation between the two would suggest that the methods yield similar predictions of which MU pairs are identical; a weak or nonlinear correlation would imply divergent predictions, complicating their correspondence.

To establish ground truth on MU identity, we generated artificial MUs from each decomposed MU by augmenting its spikes while controlling RoA at nine levels (10% increments from 10% to 90%), then computed XCC between each artificial MU and its reference (see Methods 4.5.1). This simulation approach circumvents the need for iEMG ground truth[52], as the identity relationship between each artificial MU and its reference is known by construction.

We implemented MU decomposition on high-density sEMG (HD-sEMG) data from two datasets: 1DoF, consisting of individual finger isometric contractions following a triangular force trajectory[33], and GRASP, a multi-day dataset of hand isometric gripping along a trapezoidal trajectory at different contraction levels[34] (see Methods 4.1). Decomposition yielded 10,429 MUs across 600 MU sets for the 1DoF dataset and 4,296 MUs across 240 MU sets for the GRASP dataset.

5.1.1 RoA and XCC are nonlinearly correlated

We begin by examining the functional form of the relationship between RoA and XCC across both datasets. Fitting XCC values at nine RoA levels using four candidate models incorporating MU-specific random effects and selecting the best fit via Bayesian Inference Criterion (BIC) (see Methods 4.5.1), we found that RoA and XCC are nonlinearly correlated, with XCC increasing as RoA increases across both datasets (Fig. 3a). This non-linear relationship holds for both datasets across both approaches to generate artificial MUs (surrogate and subset spike; see Methods 4.5 for details), indicating a significant population-level effect of RoA on XCC ($p < .001$; see Supplementary Note 1, see Method 4.5.1), confirming that the relationship is robust to the simulation method.

Importantly, the correspondence between RoA and XCC exhibits a dataset-dependent pattern (Fig. 3b). For the surrogate spike approach, XCC values from the 1DoF and GRASP datasets diverge significantly when RoA $\geq 50\%$, though effect sizes remain small across all levels (Cohen's r, $|r| \leq 0.1$; see Supplementary Note 2). For the subset spike approach, significant differences between datasets were observed across all nine RoA levels, again with small effect sizes (Cohen's r, $|r| \leq 0.1$; see Supplementary Note 2). These findings indicate that while the nonlinear form of the RoA-XCC relationship is consistent across datasets, its precise parameters are dataset-dependent, with direct implications for the use of universal thresholds across studies.

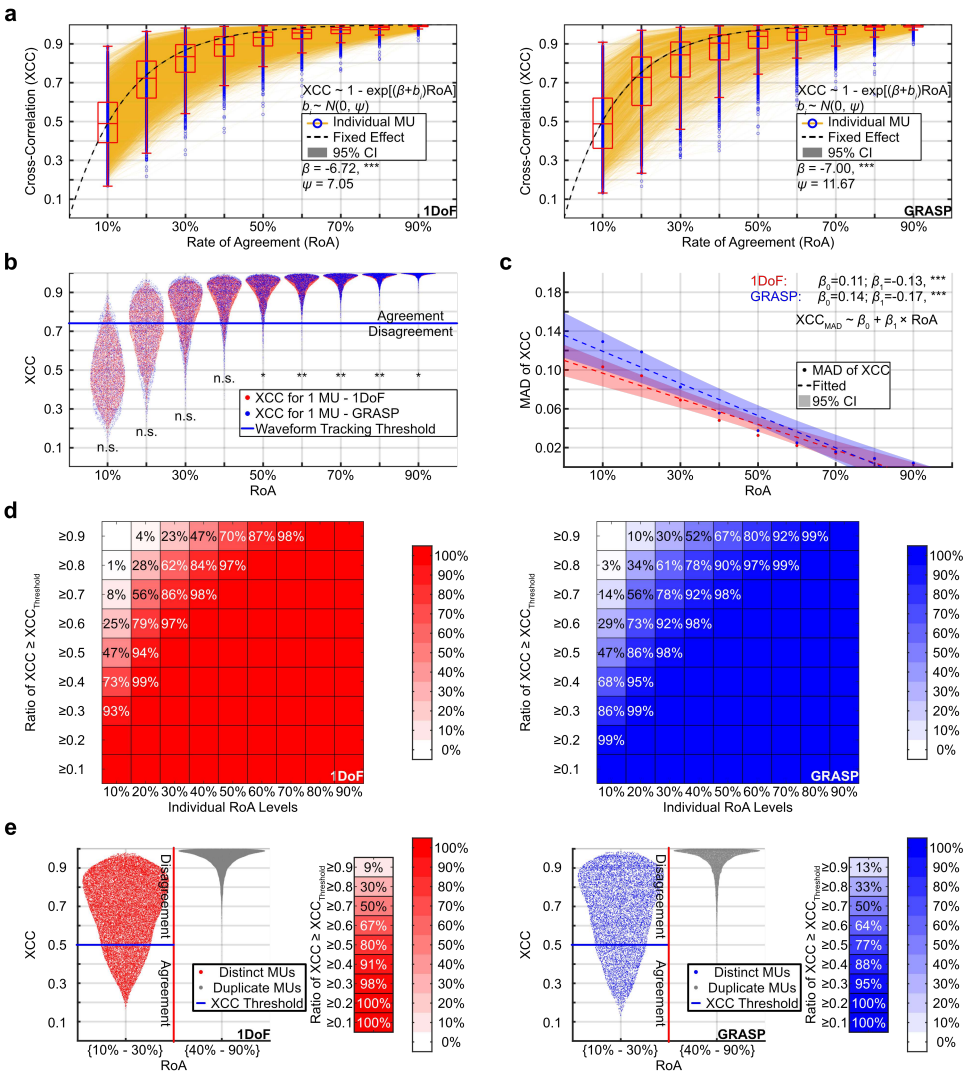


Fig. 3 RoA and XCC are nonlinearly correlated, with dataset-dependent agreement that persists after duplicate removal (surrogate spike approach). a Relationship between the spike train matching metric (RoA) and the waveform tracking metric (XCC), incorporating MU-specific random effects. An exponential mixed-effects model provided the best fit in both datasets, with RoA exerting a significant population-level effect on XCC ($p < .001$). Left: 1DoF dataset; Right: GRASP dataset. **b** Distribution of XCC values at each controlled RoA level, with horizontal lines indicating commonly used waveform-tracking thresholds. Data points above a given threshold indicate agreement between waveform tracking and spike train matching; data points below indicate disagreement. XCC variance is substantially higher at low RoA levels in both datasets. **c** Mean absolute deviation (MAD) of XCC at each RoA level, quantifying the decrease in XCC variability as RoA increases. A significant negative linear relationship between MAD and RoA was found in both datasets ($p < .001$). **d** Agreement between spike train matching and waveform tracking at each individual RoA level, computed as the proportion of MU pairs between the reference MU and the artificial MU for which both methods agree, given the applied waveform-tracking threshold. Agreement varies substantially across RoA levels and across threshold values in both datasets. Left: 1DoF dataset; Right: GRASP dataset. **e** Edge-level disagreement recomputed after excluding MU pairs with RoA > 30% (duplicate MUs[15], shown as gray dots), retaining only pairs considered distinct by spike train matching. Disagreement between the two methods persists across all commonly used waveform-tracking thresholds, even after duplicate removal. ***: $p < .001$; **: $p < .01$; *: $p < .05$.

5.1.2 XCC variability decreases with increasing RoA

Having established the nonlinear relationship between RoA and XCC, we next examined how the variability of XCC changes across RoA levels. We found that XCC exhibits substantially higher variability at low RoA levels, which decreases monotonically as RoA increases (Fig. 3b, c). To quantify this trend, we computed the mean absolute deviation (MAD) of XCC at each RoA level and performed a linear regression of MAD on RoA. A significant negative relationship was found in both datasets and both simulation approaches (surrogate spike approach: 1DoF: $R^2 = 0.944$, $\beta = -0.13$, $p < .001$; GRASP: $R^2 = 0.928$, $\beta = -0.17$, $p < .001$; subset spike approach: 1DoF: $R^2 = 0.934$, $\beta = -0.13$, $p < .001$; GRASP: $R^2 = 0.906$, $\beta = -0.14$, $p < .001$). The high XCC variance at low RoA levels, particularly below 30%, means that a given XCC value carries substantially less discriminative information when RoA is low.

5.1.3 Agreement varies at different RoA levels

Given that both XCC and its variability change significantly across RoA levels, a globally applied waveform-tracking threshold cannot equally reflect the correspondence between the two methods at all RoA levels. To quantify this, we calculated the ratio of MU pairs for which both methods agree at each RoA level, across a range of waveform-tracking thresholds (Fig. 3b, d; see Methods 4.5.2). More specifically, the agreement assumption is that if the XCC of a pair of MUs is smaller than a certain threshold (e.g., XCC $\leq$0.8), these pairs of MUs are considered to be different from each other, regardless of the tracking method used.

We found that agreement varies substantially across RoA levels for both datasets and both simulation approaches (Fig. 3d; Supplementary Fig. **??**d). At high RoA (e.g., 90%), a high XCC threshold (e.g., 0.9) guarantees complete agreement between the two methods. However, this pattern breaks down at lower RoA levels: at 10% RoA, XCC ranges from 0.1 to 1.0, making disagreement inevitable under any commonly used waveform-tracking threshold. These results demonstrate that the applicability of any fixed XCC threshold is fundamentally RoA-dependent, and that reporting RoA alongside waveform-tracking results is essential for interpreting the degree of agreement.

5.1.4 Disagreement is inevitable under any commonly used waveform tracking threshold

Even after applying the standard practice of removing duplicate MUs, defined as those with RoA exceeding 30% relative to a reference MU[15], persistent disagreement between the two methods remains under all commonly used waveform-tracking thresholds. Restricting analysis to MU pairs with RoA $\leq$ 30% and applying XCC thresholds of 0.7 and 0.9, disagreement dropped from 50% to 9% (1DoF) and from 50% to 13% (GRASP), respectively (Fig. 3e). This result demonstrates that spike train matching and waveform tracking cannot be used interchangeably, even under standard practice conditions, thereby establishing the need for a systematic, data-driven framework to characterize their correspondence.

In both simulation approaches, no single XCC threshold achieves complete agreement across the MU pool, except at the degenerate value of 0.1, which is inconsistent with reliability requirements[8,12]. This further underscores the inadequacy of universal threshold selection. Validation using the subset spike approach confirmed that all findings from the surrogate spike approach hold under this alternative simulation procedure (Supplementary Note 3; Supplementary Fig. **??**).

5.2 Distribution analysis and agreement calculation between RoA and XCC using decomposed MUs

The simulation results established a nonlinear RoA-XCC relationship with dataset-dependent characteristics and demonstrated that agreement between the two methods is RoA-level dependent. However, the simulation was restricted to nine discrete RoA levels and therefore cannot represent the full distribution of RoA values arising from actual decomposition. To address this, we performed exhaustive pairwise comparisons of all decomposed MUs within each MU set in both datasets, simultaneously computing RoA via spike train matching and XCC via waveform tracking for every pair (see Methods 4.5.3).

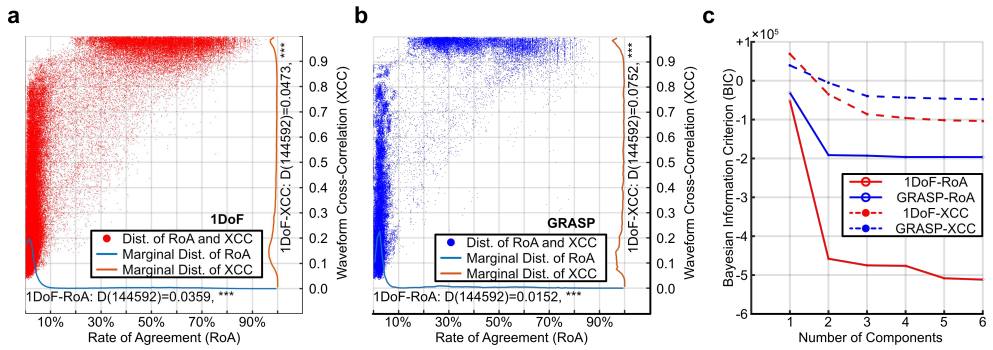


Fig. 4 Joint and marginal distributions of RoA and XCC reveal dataset-dependent patterns and non-unimodal structure across all decomposed MU sets. a, b Overlaid joint (dots) and marginal (curves) distribution of RoA and XCC for the 1DoF (a, red dots) and GRASP (b, blue dots) datasets. For each MU pair, RoA was computed via spike train matching and XCC via waveform tracking. Pairwise comparisons were performed independently within each decomposed MU set and overlaid across all sets. The two datasets exhibit a visually distinct joint distribution, consistent with the dataset-dependent patterns reported in the Results. Marginal distributions of RoA (light blue curve, bin size 1% RoA) and XCC (orange curve, bin size 0.01 XCC) for both datasets exhibit a clear non-unimodal separation between identical and distinct MU pairs. As confirmed by the Hartigan dip test ($p < .001$ for all four distributions). **c** BIC values from Gaussian Mixture Models (GMMs) fitting of the four marginal distributions in **a** and **b**, with the number of GMM components ranging from 1 to 6. The BIC decreases monotonically with increasing component number for all distributions, indicating that no stable optimal modal size can be identified. Red solid line: RoA, 1DoF dataset. Blue solid line: RoA, GRASP dataset. Red dashed line: XCC, 1DoF dataset. Blue dashed line: XCC, GRASP dataset.

5.2.1 RoA and XCC distribution reveals a dataset-dependent pattern

Extending the pairwise comparison to the full set of decomposed MUs, the joint distributions of RoA and XCC exhibit a clear dataset-dependent pattern across both datasets (Fig. 4a, b). To confirm whether the joint distribution of RoA and XCC is dataset-dependent, we first examine whether the marginal distribution of RoA and XCC shows a dataset-dependent pattern, then evaluate whether the correlation between RoA and XCC shows a dataset-dependent pattern (see Methods 4.5.3). We found the marginal distributions of both RoA and XCC differ significantly between datasets: the RoA distribution of 1DoF (Median $\pm$ MAD: $2.65 \times 10^{-2} \pm 0.02$, $n = 144,592$) differs from that of GRASP ($2.85 \times 10^{-2} \pm 0.01$, $n = 57,712$) (Wilcoxon rank-sum: $W = 1.43 \times 10^{10}$, $p = 5.78 \times 10^{-125}$), and similarly for XCC (1DoF: $2.72 \times 10^{-1} \pm 0.14$; GRASP: $2.78 \times 10^{-1} \pm 0.18$; $W = 1.47 \times 10^{10}$, $p = 4.00 \times 10^{-3}$). Effect sizes are small for both comparisons (RoA: $r = -5.29 \times 10^{-2}$; XCC: $r = 6.40 \times 10^{-3}$). These results indicate that the marginal distributions of both RoA and XCC are dataset-dependent. Furthermore, the RoA and XCC are positively correlated in both datasets (Pearson r: 1DoF $\rho = 0.82$; GRASP $\rho = 0.78$), and this correlation differs significantly between datasets (z-test: $z = 21.60$, $p < .001$). This result reveals that the joint distribution also exhibits a dataset-dependent pattern. Together, we confirm the consistent direction of dataset-dependent differences across both metrics and that both simulation and decomposed agree with each other, reinforcing the conclusion that universal thresholds applied across datasets are inappropriate.

5.2.2 RoA and XCC distributions are non-unimodal

Examining the marginal distributions more closely, both RoA and XCC are non-unimodal across both datasets (Hartigan dip test: 1DoF-RoA $D = 0.04$, $p < .001$; 1DoF-XCC $D = 0.05$, $p < .001$; GRASP-RoA $D = 0.02$, $p < .001$; GRASP-XCC $D = 0.08$, $p < .001$; Fig. 4a, b). To further estimate the number of modes, we fit Gaussian Mixture Models (GMMs) with 1 to 6 components and selected the optimal number using BIC (see Methods 4.5.3). The BIC decreased monotonically as the number of components increased for all four distributions (Fig. 4c), indicating that no single modal structure dominates (see Supplementary Note 4). This result has a direct practical implication: because neither RoA nor XCC exhibits a clear bimodal separation between same and different MU pairs, threshold values cannot be reliably determined from distributional analysis alone[7], making disagreement between the two methods inevitable and further motivating the need for dataset-specific calibration.

5.2.3 Edge-level agreement is threshold- and dataset-dependent

Having characterized the joint and marginal distributions of RoA and XCC, we now quantify how the choice of threshold affects the agreement between the two methods at the level of individual MU pairs. Edge-level agreement, defined as the ratio of pairwise MU comparisons on which both spike train matching and waveform tracking agree, was evaluated across all threshold combinations, with both RoA and XCC

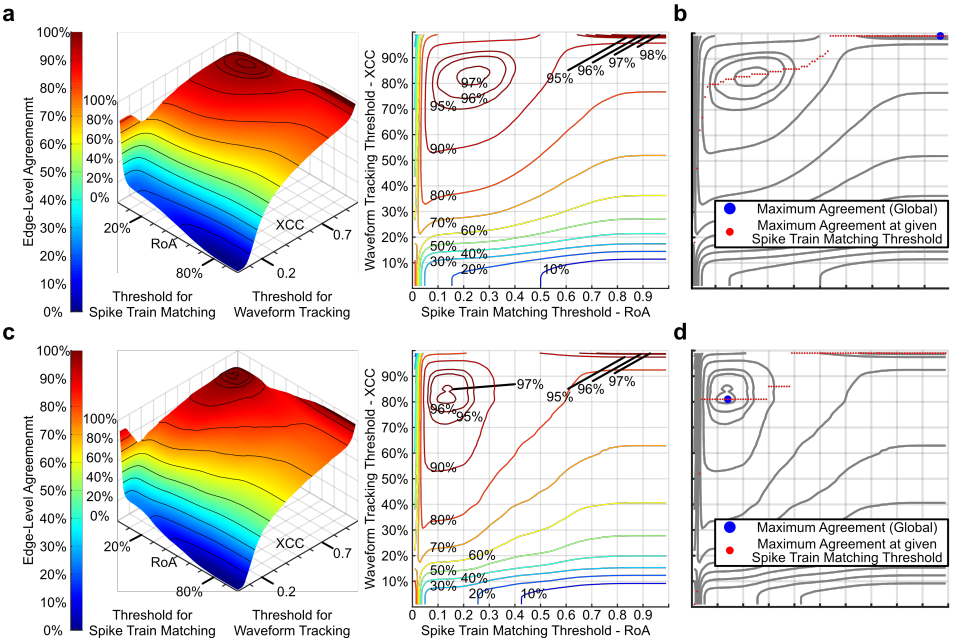


Fig. 5 Edge-level agreement between spike train matching and waveform tracking is threshold-dependent and dataset-dependent. Edge-level agreement was defined as the proportion of pairwise MU comparisons on which both methods agree (see Methods 4.5.4), and was evaluated across all combinations of spike train matching threshold (RoA, x-axis) and waveform tracking threshold (XCC, y-axis), each searched from 0.01 to 0.99 in increments of 0.01. Contour lines are drawn at 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, and 100%. **a, c** Surface (left) and contour plot (right) of edge-level agreement for the 1DoF dataset (**a**) and GRASP dataset (**c**). Each point on the surface represents the edge-level agreement achieved at a specific threshold combination. The agreement landscape differs substantially between datasets, confirming a dataset-dependent pattern. **b, d** Contour plots (gray) of edge-level agreement and threshold combinations (red and blue dots) for the 1DoF dataset (**b**) and GRASP dataset (**d**). Blue dot: the globally optimal threshold combination, achieving the highest edge-level agreement across all possible threshold combinations (1DoF: RoA = 96%, XCC = 0.99, agreement = 98.60%; GRASP: RoA = 14%, XCC = 0.81, agreement = 97.21%). Neither optimal combination corresponds to conventionally used threshold values. Red dots: the waveform tracking threshold yielding maximum edge-level agreement at each fixed spike train matching threshold, illustrating the dataset-dependent trajectory of optimal threshold pairs across the full parameter space.

thresholds searched from 0.01 to 0.99 in increments of 0.01 (Fig. 5; see Methods 4.5.4). Three key findings emerge.

First, the maximum achievable agreement is dataset-dependent and falls below 100% in both approaches, confirming that disagreement between the two methods is irreducible. For the 1DoF dataset, the globally optimal threshold combination (RoA = 96%, XCC = 0.99) achieves a maximum edge-level agreement of 98.60% (Fig. 5b, blue dot). However, such high levels of thresholds on RoA and XCC can result in the interpretation that all the decomposed MUs are distinct and no MU could be identified as the same if these thresholds are used across trials (see Discussion 6.5 for full elaboration). For the GRASP dataset, the globally optimal combination (RoA =

14%, XCC = 0.81) achieves 97.21% (Fig. 5d, blue dot). Neither optimal combination aligns with conventionally used thresholds (RoA = 30%, XCC = 0.7-0.9).

Second, fixing the spike train matching threshold at the commonly used value of 30% RoA[15,19,21] and optimizing over the waveform-tracking threshold yields a maximum edge-level agreement of 97.02% at XCC = 0.84 for 1DoF and 83.65% at XCC = 0.86 for GRASP. Applying the conventional waveform tracking threshold within the range of 0.7-0.9 at 30% RoA ensures at least 90% pairwise agreement for 1DoF and 80% for GRASP, but with optimal thresholds that differ between datasets and can deviate from standard practice (i.e., 0.7, 0.8, 0.9).

Third, the waveform-tracking threshold that maximizes edge-level agreement at each fixed spike train matching threshold follows a dataset-dependent trajectory across the full parameter space (Fig. 5b, d, red dots; see Supplementary Note 5). Taken together, these results demonstrate that the correspondence between the two methods is fundamentally sensitive to threshold selection and cannot be adequately characterized by a single universal threshold combination.

5.2.4 Node-level agreement is threshold- and dataset-dependent

Edge-level agreement captures pairwise consistency but does not directly reflect how the two methods agree in assigning individual MUs to groups. To address this, we define node-level agreement as the proportion of MUs clustered into the same group by both methods, quantified using the symmetric purity score[47] (see Methods 4.5.5). Thresholds were searched from 0.05 to 0.95 in increments of 0.05, and the median symmetric purity score was used to represent node-level agreement at each threshold combination (Fig. 6).

Node-level agreement exhibits a dataset-dependent pattern consistent with edge-level findings. For the 1DoF dataset, the globally optimal threshold combination (RoA = 25%, XCC = 0.80) achieves a maximum node-level agreement of 96.30% (Fig. 6b, blue dots). For the GRASP dataset, complete node-level agreement (100%) is achieved across a range of threshold combinations (RoA = 10-20%, XCC = 0.8-0.9; Fig. 6d, blue dot). Fixing spike train matching at 30% RoA, the waveform-tracking threshold that maximizes node-level agreement is XCC = 0.80 for 1DoF, achieving 95% agreement, and XCC = 0.85 for GRASP, achieving 78.17%. Both optimal sets of threshold combinations deviate from the conventionally used range and differ between datasets. The full trajectory of optimal waveform-tracking thresholds across all spike train matching thresholds follows a complex, dataset-dependent pattern (Fig. 6, red dots; see Supplementary Note 5).

5.2.5 Group-level agreement is threshold- and dataset-dependent

Node-level agreement may be inflated by the presence of duplicate MUs, since a decomposition containing many duplicates can still yield high clustering consistency. To provide an assessment that is robust to this concern, we introduce group-level agreement, quantified using the macro Jaccard index[48–50], which treats each unique MU group equally and is therefore insensitive to the number of duplicate MUs within a group (see Methods 4.5.6). Thresholds were searched from 0.05 to 0.95 in increments of

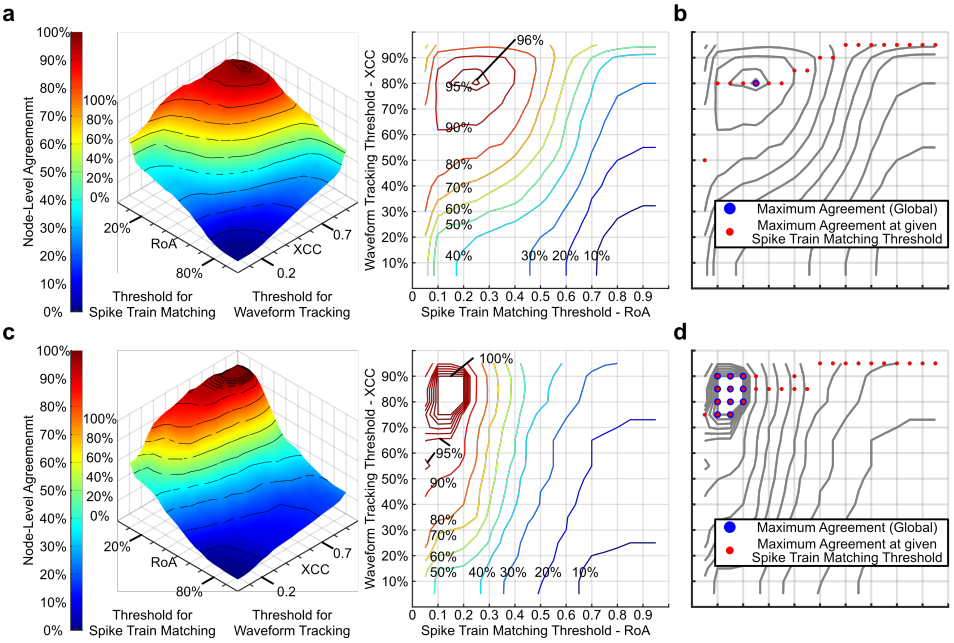


Fig. 6 Node-level agreement between spike train matching and waveform tracking is threshold-dependent and dataset-dependent. Node-level agreement was defined as the proportion of MUs clustered into the same group by both methods, quantified using the symmetric purity score (see Methods 4.5.5). The median symmetric purity score across all decomposed MU sets was used to represent node-level agreement at each threshold combination, and was evaluated across all combinations of spike train matching threshold (RoA, x-axis) and waveform tracking threshold (XCC, y-axis), each searched from 0.05 to 0.95 in increments of 0.05. Contour lines are drawn at 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, and 100%. **a, c** Surface (left) and contour plot (right) of node-level agreement for the 1DoF dataset (**a**) and GRASP dataset (**c**). Each point on the surface represents the node-level agreement achieved at a specific threshold combination. The agreement landscape differs substantially between datasets, confirming a dataset-dependent pattern. **b,d** Contour plots (gray) of node-level agreement and threshold combinations (red and blue dots) for the 1DoF (**b**) and GRASP (**d**) dataset. Blue dots: the globally optimal threshold combinations, achieving the highest node-level agreement across all possible threshold combinations (1DoF: RoA = 25%, XCC = 0.80, agreement = 96.30%; GRASP: RoA = 10–20%, XCC = 0.80–0.90, agreement = 100%). Notice that for the GRASP dataset, a complete node-level agreement (100%) can be achieved. Neither globally optimal combination in either dataset corresponds to the conventionally used threshold values (RoA = 30%, XCC = 0.7–0.9). Red dots: the waveform tracking threshold yielding maximum node-level agreement at each fixed spike train matching threshold. The optimal threshold trajectory differs substantially from that of the 1DoF dataset, confirming a dataset-dependent pattern.

0.05, and the median macro Jaccard index was used to represent group-level agreement at each threshold combination (Fig. 7).

Group-level agreement mirrors the dataset-dependent pattern observed at the edge and node levels. For the 1DoF dataset, maximum group-level agreement of 97.69% is achieved at RoA = 25%, XCC = 0.80 (Fig. 7b, blue dot), indicating that on average, a unique MU group identified by spike train matching at 25% RoA overlaps 97.69% with the corresponding group identified by waveform tracking at XCC = 0.80. For the

GRASP dataset, complete group-level agreement (100%) is achieved across RoA = 10-15% and XCC = 0.75-0.90 (Fig. 7d, blue dots). Fixing spike train matching at 30% RoA, maximum group-level agreement is 95% at XCC = 0.80 for 1DoF and 78.17% at XCC = 0.90 for GRASP. The optimal waveform-tracking threshold at each fixed spike train matching threshold again follows a complex, dataset-dependent trajectory (Fig. 7, red dots; see Supplementary Note 5).

Across all three levels of analysis, edge, node, and group, the agreement between spike train matching and waveform tracking is consistently threshold-dependent and dataset-dependent, with optimal threshold combinations that deviate from conventional values and vary between datasets. These convergent findings across three independent graph-level measures provide strong evidence that the use of fixed, universal thresholds is inappropriate and that dataset-specific pilot studies are necessary before deploying either method.

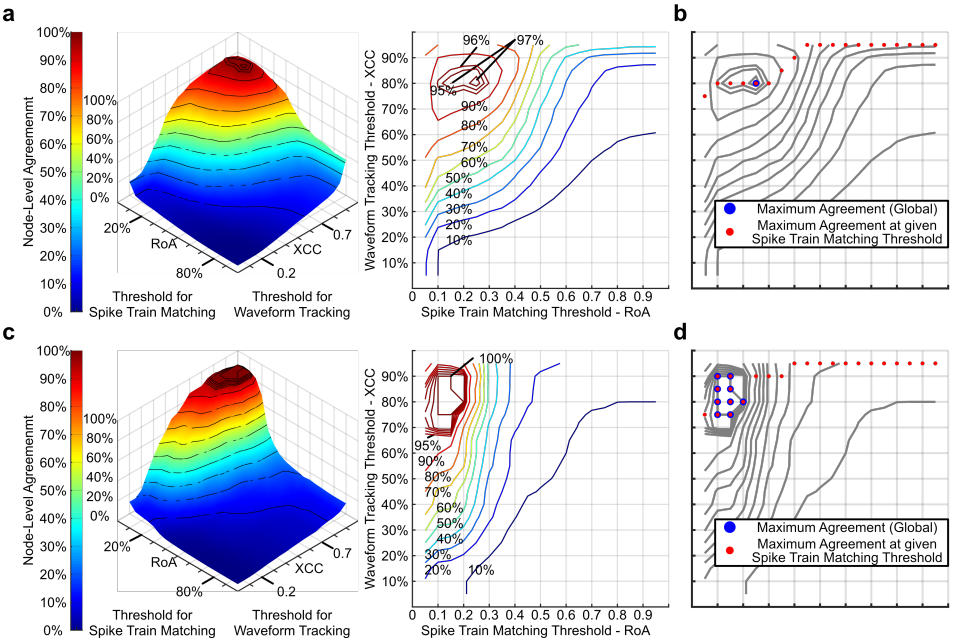


Fig. 7 Group-level agreement between spike train matching and waveform tracking is threshold-dependent and dataset-dependent. Group-level agreement was defined as the average overlap between unique MU groups identified by the two methods, quantified using the macro Jaccard index (see Methods 4.5.6). Unlike node-level agreement, the macro Jaccard index treats each unique MU group equally and is therefore insensitive to the number of duplicate MUs within a group, providing a more conservative and physiologically direct measure of correspondence. The median macro Jaccard index across all decomposed MU sets was used to represent group-level agreement at each threshold combination, and was evaluated across all combinations of spike train matching threshold (RoA, x-axis) and waveform tracking threshold (XCC, y-axis), each searched from 0.05 to 0.95 in increments of 0.05. Contour lines are drawn at 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, and 100%. **a, c** Surface (left) and contour plot (right) of group-level agreement for the 1DoF (**a**) and GRASP (**c**) dataset. Each point on the surface represents the group-level agreement achieved at a specific threshold combination. The agreement landscape differs substantially between datasets, confirming a dataset-dependent pattern. **b, d** Contour plots (gray) of group-level agreement and threshold combinations (red and blue dots) for the 1DoF (**b**) and GRASP (**d**) dataset. Blue dots: the globally optimal threshold combinations, achieving the highest group-level agreement across all possible threshold combinations (1DoF: RoA = 25%, XCC = 0.80, agreement = 97.69%; GRASP: RoA = 10–15%, XCC = 0.75–0.90, agreement = 100%). Notice that for the GRASP dataset, a complete group-level agreement (100%) can be achieved. Neither globally optimal combination in either dataset corresponds to the conventionally used threshold values (RoA = 30%, XCC = 0.7–0.9). Red dots: the waveform tracking threshold yielding maximum group-level agreement at each fixed spike train matching threshold. The optimal threshold trajectory differs substantially from that of the 1DoF dataset, confirming a dataset-dependent pattern.

6 Discussion

In this study, we examined how spike train matching and waveform tracking agree in identifying the same MUs by characterizing the relationship between their respective metrics and comparing their outputs at three levels of structural resolution, using both simulated and real decomposed MU data from two datasets involving isometric contractions with different force trajectories. Despite the widespread use of both methods across neuroscience, kinesiology, and rehabilitation research[8,11,13–16], their correspondence has never been systematically examined. Our results reveal that the two methods are not interchangeable, that their agreement is both threshold-dependent and dataset-dependent, and that commonly used threshold values require careful reconsideration.

6.1 A high MUAP waveform XCC alone is not sufficient for identifying the same MU

A central assumption in waveform tracking is that a high XCC reliably indicates that two MUs are identical[8,12]. Our findings challenge this assumption directly. Using simulated data, we showed that XCC increases nonlinearly with RoA, but with substantially higher variability at low RoA levels, a pattern that persists regardless of the approach used to generate artificial MUs. This means that a high XCC value can occur even when two MUs have very low spike train overlap (low RoA), simply because of the large spread of XCC values at low RoA. The same pattern was confirmed in the real decomposed MUs: at RoA below 10%, XCC values spanning from 0.1 to 0.9 were observed in both datasets, and some MU pairs reached XCC $\geq$ 0.9 despite RoA below 10% (Fig. 4; Supplementary Fig. **??**).

A concrete example from the 1DoF dataset illustrates this problem. Two MUs, decomposed from the same trial of HD-sEMG recording, showed clear disagreement between spike train matching and waveform tracking: they would be identified as the same MU under waveform tracking but as different MUs under spike train matching. Inspection of additional MU characteristics (see Methods 4.2.3; Supplementary Fig. **??**) suggests that the peak-to-peak value of the MUAP is different. Considering that the peak-to-peak value of the MUAP is often used to estimate MU position[43] or MU territory[44], this discrepancy may be due to different MUs occupying the same skin location but at different depths. The mutually exclusive discharge behavior of motor neurons reported by Zhu et al.[53] provides a physiological basis for why two spatially distinct MUs could produce similar MUAP waveforms while firing at non-overlapping times. Our observation is also consistent with the description[54] that "MUAPs from different MUs may have similar amplitude and shape when the muscle fibers of each MU in the detectable vicinity of the electrode have a similar spatial arrangement".

These findings have direct practical implications. When two MUs are decomposed from the same contraction and waveform tracking is used to assess their identity, researchers should also report the RoA between their MUSTs. If spike train data are unavailable, reporting the MUAP peak-to-peak voltage and the estimated muscle depth can serve as additional discriminating criteria to guard against false identifications driven by high XCC at low RoA. More broadly, the assumption that a higher

waveform tracking threshold guarantees more reliable MU identification[12] is valid only when RoA is sufficiently high. At low RoA, the high variability of XCC undermines this guarantee, and XCC alone cannot be treated as a sufficient criterion for MU identity. This calls for greater caution among researchers and neuroscientists when reporting MU identification results based on waveform tracking alone.

Our findings also add new context to prior work by del Vecchio et al.[55], who showed that using only the first or last 20 spikes of a MUST can still yield an average waveform correlation of approximately 0.86. Based on our simulation results, a subset of 20 spikes from a full MUST implies a low effective RoA relative to the complete train, and the wide spread of XCC values at low RoA levels means that such high average XCC values may mask substantial variability at the individual MU pair level. Differences in muscle (forearm muscles FDS and EDC in our study vs. tibialis anterior in theirs), spatial filter choice (Linear Single Differential in our work vs. Double Differential in theirs), and XCC implementation (1D vs. 2D[56]) may further contribute to this discrepancy. Farina et al.[40] showed that second-order filters (IR and LDD) produce significantly shorter MU action potential durations than LSD, which would affect XCC values and thereby alter the apparent threshold at which waveform tracking is reliable.

6.2 Dataset-dependent patterns challenge the use of universal thresholds

A consistent dataset-dependent pattern was observed across all analyses, regardless of whether simulated or real decomposed MUs were used, and regardless of whether RoA or XCC was the metric under examination. This pattern was further confirmed at all three levels of agreement measurement using graph analysis: edge, node, and group. The practical implication is direct: threshold values optimized for one dataset cannot be assumed to generalize to another, even within the same broad category of isometric contractions.

This finding directly challenges current practice. A waveform-tracking threshold of 0.8 has been applied to datasets involving isometric contraction[13] and to datasets involving concentric and eccentric contraction[14], without examining whether the same threshold is appropriate across these different task conditions. Our results provide quantitative evidence that this approach will introduce false-positive pairs after the MU identification if researchers consider spike train matching as the ground truth. Within the isometric contraction category alone, the 1DoF dataset (triangular force trajectory) and the GRASP dataset (trapezoidal force trajectory) yielded different optimal threshold combinations at all three agreement levels. This task-dependence is consistent with evidence that non-trapezoidal contraction protocols can produce recruitment patterns that violate the orderly recruitment observed in primate studies[27], while trapezoidal contractions are more likely to respect the Henneman size principle[57]. The difference in recruitment dynamics between the two tasks likely contributes to differences in MU pool composition and, consequently, to the different RoA and XCC distributions observed across datasets.

The non-unimodal but modally indeterminate distributions of both RoA and XCC (Supplementary Note 4) further reinforce this point. Several studies have proposed determining thresholds analytically from the distributional properties of these metrics,

under the assumption that a bimodal distribution would reveal a natural boundary between same and different MU pairs[7]. Chen and Zhou[7] observed a bimodal distribution in two-source validation results comparing sEMG and iEMG, and proposed a matching threshold as high as 81.82% RoA on that basis. Our distributions, derived from pairwise comparisons within sEMG decomposition alone, show no clear bimodal structure and no stable optimal modal size as estimated by BIC. This discrepancy is likely attributable to the different comparison context: two-source validation contrasts MUs across recording modalities, while our analysis operates entirely within sEMG decomposition. It may also reflect the inter-subject variability introduced by overlaying pairwise comparisons across subjects, as cautioned by De Luca et al.[58]. Together, these results suggest that threshold determination from distributional analysis alone is not reliably applicable in the sEMG decomposition context, and that dataset-specific pilot studies are necessary.

6.3 Spike train matching and waveform tracking can achieve high agreement but remain non-interchangeable

Despite the dataset-dependent nature of the optimal thresholds, both methods can achieve high agreement under the right conditions. For the 1DoF dataset, an edge-level agreement of 98.60% is achievable at (RoA = 96%, XCC = 0.99), and a node-level agreement of 96.30% at (RoA = 25%, XCC = 0.80). For the GRASP dataset, complete node- and group-level agreement (100%) is achievable across a range of threshold combinations (RoA = 10-20%, XCC = 0.75-0.90). However, these optimal combinations differ substantially between datasets and do not align with any single pair of conventionally used thresholds, confirming that high agreement is achievable but only with dataset-specific calibration.

Critically, even the maximum achievable agreement falls below 100% at the edge level in both datasets, confirming that some degree of disagreement between the two methods is irreducible. This irreducibility follows directly from the nonlinear RoA-XCC relationship and the high XCC variability at low RoA: there will always be MU pairs for which the two methods reach different conclusions, regardless of how thresholds are set. This is not a measurement artifact but rather reflects a fundamental difference in what the two metrics capture. RoA measures the temporal overlap of spike events, while XCC measures the spatial similarity of MUAP waveforms. These are related but not equivalent properties of a MU pair. It is physiologically plausible for them to disagree, particularly for MUs that share similar muscle fiber territory (producing similar waveforms) but fire at non-overlapping times (producing low spike overlap).

The non-interchangeability of the two methods also has implications for how duplicate removal is implemented prior to waveform tracking. The standard approach to duplicate removal[15,19] sorts MUSTs by discharge count and performs recursive pairwise RoA comparisons, without constructing the full pairwise comparison graph. Our graph-based approach, consistent with Formento et al.[32], performs exhaustive pairwise comparisons and preserves the full connectivity structure. As shown in our earlier work[59], these two approaches can yield discrepant results. Researchers who combine

spike train matching for duplicate removal with waveform tracking for cross-trial identification should be aware that the graph structure underlying duplicate removal may influence the subsequent tracking results in ways that are not captured by standard implementations.

6.4 Agreement challenges the use of commonly accepted threshold values

Our agreement analysis provides a systematic, quantitative basis for evaluating the threshold values currently used in both spike train matching and waveform tracking. Using 30% RoA as the spike train matching reference, consistent with common practice[15,16,19,21], applying a waveform-tracking threshold between 0.7 and 0.9 ensures at least 90% edge-level agreement for 1DoF and at least 80% for GRASP. At the node level, the same operation ensures that at least 90% of MUs (1DoF) and 70% of MUs (GRASP) are clustered into the same groups. At the group level, at least 80% (1DoF) and 40% (GRASP) of unique MU identities overlap between the two methods. The decreasing agreement from edge to node to group level is expected: edge-level agreement reflects pairwise consistency, while group-level agreement is the most stringent criterion, requiring complete overlap of identified MU communities. For researchers seeking a rapid preliminary assessment, edge-level calculation offers a practical first step, while group-level agreement provides the most physiologically meaningful and conservative estimate.

The globally optimal threshold combinations, which do not include 30% RoA in either dataset, raise questions about the physiological basis of this widely used value. The 30% threshold was introduced in Holobar et al.[15] without explicit physiological validation and has since been adopted broadly without further scrutiny[16,19–21]. For the GRASP dataset, the globally optimal edge-level combination (RoA = 14%, XCC = 0.81) implies that if a waveform-tracking threshold of 0.81 is considered physiologically valid, then a spike train matching threshold of 14% may be equally justified on agreement grounds, which directly challenges the conventional 30% value. Whether thresholds as low as 14% or as high as 96% are physiologically meaningful requires dedicated investigation that goes beyond the scope of agreement analysis alone.

The threshold-based framework itself also warrants scrutiny. Both spike train matching and waveform tracking reduce a continuous pairwise similarity measure to a binary same-or-different decision by applying a single threshold, implicitly assuming that MUs form two distinct populations: identical and non-identical. However, empirical evidence from primate[28] and human studies[60–62] indicates that MUST firing patterns and motor unit modes exhibit more than two distinct behavioral clusters. A richer framework that accommodates graded similarity and multi-modal MU population structure, for example, through community detection in weighted graphs as explored here, may ultimately provide a more physiologically faithful representation than threshold-based bisection alone. This is consistent with the broader view of motor unit organization articulated by Duchateau and Enoka[63], in which the integrated behavior of the motor unit pool reflects a continuum of neural inputs rather than a binary classification.

Finally, when using a 30% spike train matching threshold as reference, our results show that both datasets achieve maximum agreement when the waveform-tracking threshold is between 0.8 and 0.9 across all three agreement levels. This means the conventional lower bound of 0.7 for waveform tracking should be used with particular caution, as it yields lower agreement with spike train matching than thresholds of 0.8 or 0.9 in both datasets. Rather than recommending a single fixed value, we provide the full agreement surface across all threshold combinations (Supplementary Note 5) to allow researchers to make informed, dataset-specific decisions.

6.5 Limitations and future directions

Several limitations of this study should be noted. First, our analysis is restricted to within-trial comparisons, since direct spike train matching across trials is not feasible because exact spike timing cannot be preserved across separate contractions[8,18]. Extending the agreement framework to cross-trial settings, where filter tracking and waveform tracking are the available options, remains an important direction for future work.

Second, the graph-based agreement analysis requires constructing pairwise comparison graphs across all decomposed MU sets. Because decomposed MU sets are defined by experimental conditions rather than physiological boundaries, and because each MU set typically contains fewer than 30 MUs due to the orthonormality constraints of blind source separation, the graph structures analyzed here are relatively sparse. Conventional community detection methods typically perform best with larger networks. Future work using datasets with higher MU yield, or combining MU sets across trials under carefully controlled conditions, would allow more robust graph-level analyses.

Third, this study uses two datasets with isometric contractions that differ in force trajectory. Extending this analysis to dynamic contractions, including concentric and eccentric tasks, would test the generalizability of the dataset-dependent patterns observed here and provide broader guidance for threshold selection across movement conditions.

Fourth, while our findings strongly suggest the need for dataset-specific pilot studies before applying fixed thresholds, we do not prescribe a specific procedure for conducting such pilot studies. Developing a principled, computationally efficient calibration protocol, potentially based on a subset of the pairwise agreement surface explored here, is an important practical challenge for future work.

Finally, although we provided thresholds that enable researchers to make informed, dataset-specific decisions, we do not consider the preserved MU counts after applying these thresholds. For example, in the 1DoF dataset, even though a threshold combination (RoA = 96%, XCC = 0.99) can achieve a globally maximum edge-level agreement of 98.60%, applying this threshold removed all 144,592 comparisons (kept all MUs) in spike train matching (RoA $\geq$ 96%). Such a high threshold removes all possible connectivity among MUs, as Chen and Zhou[7] noted, leaving no same MU pairs available that can be used for further study. In fact, more MUs will be preserved rather than focusing on the highest agreement; choosing a threshold combination of (RoA = 24%, XCC = 0.84) will achieve a 97.26% edge-level agreement. while only removing 117,634 comparisons, leaving more of the same MU pairs that can be used

for further study. Overall, a trade-off exists between the number of MU after identification and the agreement accuracy. Researchers should therefore balance agreement accuracy with MU retention based on the specific goals of their analysis, rather than prioritizing maximal agreement alone.

7 Conclusion

This study provides the first systematic examination of the correspondence between spike train matching and waveform tracking, two widely used methods for identifying the same MUs from sEMG recordings. Using both simulated and real decomposed MUs from two datasets involving isometric contractions with different force trajectories, and evaluating agreement at three levels of graph-theoretic resolution, we draw the following conclusions.

First, RoA and XCC are nonlinearly correlated, with XCC increasing as RoA increases. However, XCC exhibits substantially higher variability at low RoA levels, and this variability decreases monotonically as RoA increases. This finding has a direct implication for practice: a high XCC value does not by itself guarantee that two MUs are identical, particularly when the underlying spike train overlap is low. Researchers should therefore exercise caution when identifying the same MUs based solely on MUAP waveform similarity, and should report RoA alongside XCC whenever both can be computed.

Second, spike train matching and waveform tracking can achieve high agreement, but they are not interchangeable. The maximum achievable agreement falls below 100% at the edge level in both datasets, confirming that some degree of disagreement between the two methods is irreducible. This irreducibility reflects a fundamental difference between what RoA and XCC measure: the temporal overlap of spike events versus the spatial similarity of MUAP waveforms. These are related but not equivalent properties, and the two methods will not always reach the same conclusion about MU identity.

Third, the agreement between the two methods is both threshold-dependent and dataset-dependent, and this pattern is consistent across all three levels of analysis: edge, node, and group. The conventionally used threshold combination of 30% RoA for spike train matching and 0.7-0.9 XCC for waveform tracking ensures at least 90% edge-level agreement and at least 90% node-level agreement for the 1DoF dataset, but only 80% edge-level and 70% node-level agreement for the GRASP dataset. At the group level, the same combination yields at least 80% agreement for 1DoF and as low as 40% for GRASP. These differences underscore that the same threshold values cannot be applied uniformly across datasets, even within the same category of isometric contractions.

Fourth, the marginal distributions of both RoA and XCC are non-unimodal across both datasets, with no clear bimodal separation that would permit threshold determination from distributional analysis alone. This rules out the analytical determination of thresholds from the data distribution and further motivates the need for dataset-specific calibration.

Taken together, these findings provide strong quantitative evidence that the use of fixed, universal thresholds for spike train matching or waveform tracking is inappropriate. A dataset-specific pilot study, following the framework demonstrated in Section 5.2, is strongly recommended prior to applying either method. The graph-based agreement framework introduced here, evaluating correspondence at the edge, node, and group levels simultaneously, offers a principled and extensible approach for characterizing the relationship between MU identification methods and may serve as a foundation for future work on cross-trial tracking, dynamic contractions, and the development of data-driven threshold calibration procedures.

8 Declarations

8.1 Ethics approval and consent to participate

Ethics approval and informed consent do not apply to this study, as all analyses were conducted using previously published, publicly available datasets.

8.2 Consent for publication

Not applicable.

8.3 Availability of data and materials

All datasets used in this study are publicly available. Readers can refer to Jiang et al.[33] and Yang et al.[34] for downloading the original datasets. The decomposed datasets are available upon request to the authors or can be reproduced using the code provided in the original studies.

All code, including data preprocessing, HD-sEMG decomposition, artificial MU generation, spike train matching, waveform tracking, cluster analysis, and statistical analysis, is available at https://github.com/LIMB-UCF/sEMG_Decomp.

8.4 Competing interests

The author MR is an Editorial Board Member of JNER.

8.5 Funding

We acknowledge the University of Central Florida Faculty Cluster Initiative Student Scholarship (Zhang).

8.6 Authors' contributions

These authors jointly supervised this work: MR and YW. SZ, MR, and YW conceived and designed the study; SZ processes all the data. SZ and YW prepare and validate the sEMG decomposition code. SZ conducts all data analysis. MR and YW validate the processed data and the analysis results. SZ, MR, and YW wrote the first draft. SZ, MR, and YW revised the paper. All authors contributed to the editing of the paper.

8.7 Acknowledgements

We express our sincere gratitude to Erin Weidel, Mae Torra, and Trevor Overton from the laboratory of machine interactions and the brain for their valuable assistance in improving the grammar of this paper.

9 References

References

[1] Finsterer, J., Erdorf, M., Mamoli, B., Fuglsang-Frederiksen, A.: Needle electromyography of bulbar muscles in patients with amyotrophic lateral sclerosis: evidence of subclinical involvement. Neurology **51**(5), 1417–1422 (1998)

[2] Nishikawa, Y., Holobar, A., Watanabe, K., Takahashi, T., Ueno, H., Maeda, N., Maruyama, H., Tanaka, S., Hyngstrom, A.S.: Detecting motor unit abnormalities in amyotrophic lateral sclerosis using high-density surface emg. Clinical Neurophysiology **142**, 262–272 (2022)

[3] Goodlich, B.I., Del Vecchio, A., Kavanagh, J.J.: Motor unit tracking using blind source separation filters and waveform cross-correlations: reliability under physiological and pharmacological conditions. Journal of Applied Physiology **135**(2), 362–374 (2023)

[4] Li, X., Holobar, A., Gazzoni, M., Merletti, R., Rymer, W.Z., Zhou, P.: Examination of poststroke alteration in motor unit firing behavior using high-density surface emg decomposition. IEEE transactions on biomedical engineering **62**(5), 1242–1252 (2014)

[5] Herda, T.J.: Resistance exercise training and the motor unit. European Journal of Applied Physiology **122**(9), 2019–2035 (2022)

[6] De Boer, E., Kuyper, P.: Triggered correlation. IEEE Transactions on Biomedical Engineering (3), 169–179 (2008)

[7] Chen, M., Zhou, P.: Caution is necessary for acceptance of motor units with intermediate matching in surface emg decomposition. Frontiers in Neuroscience **16**, 876659 (2022)

[8] Martinez-Valdes, E., Negro, F., Laine, C., Falla, D., Mayer, F., Farina, D.: Tracking motor units longitudinally across experimental sessions with high-density surface electromyography. The Journal of physiology **595**(5), 1479–1496 (2017)

[9] Oßwald, M., Cakici, A.L., Oliveira, D., Braun, D.I., Farina, D., Del Vecchio, A.: Task-specific motor units in the extrinsic hand muscles control single-and multidigit tasks of the human hand. Journal of Applied Physiology **138**(5), 1187–1200 (2025)

[10] Chen, C., Zhu, X.: Tracking motor unit discharges across days based on the pre-trained separation vectors. In: 2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob), pp. 681–686 (2024). IEEE

[11] Lulic-Kuryllo, T., Thompson, C.K., Jiang, N., Negro, F., Dickerson, C.R.: Neural control of the healthy pectoralis major from low-to-moderate isometric contractions. Journal of Neurophysiology **126**(1), 213–226 (2021)

[12] Valli, G., Ritsche, P., Casolo, A., Negro, F., De Vito, G.: Tutorial: Analysis of central and peripheral motor unit properties from decomposed high-density surface emg signals with openhdemg. Journal of Electromyography and Kinesiology **74**, 102850 (2024)

[13] Cudicio, A., Martinez-Valdes, E., Cogliati, M., Orizio, C., Negro, F.: The force-generation capacity of the tibialis anterior muscle at different muscle–tendon lengths depends on its motor unit contractile properties. European Journal of Applied Physiology **122**(2), 317–330 (2022)

[14] Oliveira, A.S., Negro, F.: Neural control of matched motor units during muscle shortening and lengthening at increasing velocities. Journal of applied physiology **130**(6), 1798–1813 (2021)

[15] Holobar, A., Minetto, M.A., Botter, A., Negro, F., Farina, D.: Experimental analysis of accuracy in the identification of motor unit spike trains from high-density surface emg. IEEE Transactions on Neural Systems and Rehabilitation Engineering **18**(3), 221–229 (2010)

[16] Negro, F., Muceli, S., Castronovo, A.M., Holobar, A., Farina, D.: Multi-channel intramuscular and surface emg decomposition by convolutive blind source separation. Journal of neural engineering **13**(2), 026027 (2016)

[17] Chen, M., Zhang, X., Lu, Z., Li, X., Zhou, P.: Two-source validation of progressive fastica peel-off for automatic surface emg decomposition in human first dorsal interosseous muscle. International journal of neural systems **28**(09), 1850019 (2018)

[18] Carroll, T., Selvanayagam, V., Riek, S., Semmler, J.: Neural adaptations to strength training: moving beyond transcranial magnetic stimulation and reflex studies. Acta physiologica **202**(2), 119–140 (2011)

[19] Avrillon, S., Hug, F., Enoka, R.M., Caillet, A.H., Farina, D.: The identification of extensive samples of motor units in human muscles reveals diverse effects of neuromodulatory inputs on the rate coding. Elife **13**, 97085 (2024)

[20] Škarabot, J., Beauchamp, J.A., Pearcey, G.E.: Human motor unit discharge patterns reveal differences in neuromodulatory and inhibitory drive to motoneurons

across contraction levels. Journal of Neurophysiology **134**(5), 1429–1444 (2025)

[21] Dernoncourt, F., Avrillon, S., Logtens, T., Cattagni, T., Farina, D., Hug, F.: Flexible control of motor units: is the multidimensionality of motor unit manifolds a sufficient condition? The Journal of Physiology **603**(8), 2349–2368 (2025)

[22] Frančič, A., Holobar, A.: On the reuse of motor unit filters in high density surface electromyograms recorded at different contraction levels. Ieee Access **9**, 115227–115236 (2021)

[23] Valli, G., Sarto, F., Casolo, A., Del Vecchio, A., Franchi, M.V., Narici, M.V., De Vito, G.: Lower limb suspension induces threshold-specific alterations of motor units properties that are reversed by active recovery. Journal of Sport and Health Science **13**(2), 264–276 (2024)

[24] Dai, C., Hu, X.: Independent component analysis based algorithms for high-density electromyogram decomposition: Systematic evaluation through simulation. Computers in biology and medicine **109**, 171–181 (2019)

[25] Casolo, A., Farina, D., Falla, D., Bazzucchi, I., Felici, F., Del Vecchio, A.: Strength training increases conduction velocity of high-threshold motor units. Medicine and Science in Sports and Exercise **52**(4), 955–967 (2020)

[26] Del Vecchio, A., Casolo, A., Negro, F., Scorcelletti, M., Bazzucchi, I., Enoka, R., Felici, F., Farina, D.: The increase in muscle force after 4 weeks of strength training is mediated by adaptations in motor unit recruitment and rate coding. The Journal of physiology **597**(7), 1873–1887 (2019)

[27] Marshall, N.J., Glaser, J.I., Trautmann, E.M., Amematsro, E.A., Perkins, S.M., Shadlen, M.N., Abbott, L., Cunningham, J.P., Churchland, M.M.: Flexible neural control of motor units. Nature neuroscience **25**(11), 1492–1504 (2022)

[28] Palmer, S.S., Fetz, E.E.: Discharge properties of primate forearm motor units during isometric muscle activity. Journal of neurophysiology **54**(5), 1178–1193 (1985)

[29] Humphries, M.D.: Spike-train communities: finding groups of similar spike trains. Journal of Neuroscience **31**(6), 2321–2336 (2011)

[30] Hug, F., Avrillon, S., Sarcher, A., Del Vecchio, A., Farina, D.: Correlation networks of spinal motor neurons that innervate lower limb muscles during a multi-joint isometric task. The Journal of physiology **601**(15), 3201–3219 (2023)

[31] Avrillon, S., Hug, F., Farina, D.: A graph-based approach to identify motor neuron synergies. BioRxiv, 2023–02 (2023)

[32] Formento, E., Botros, P., Carmena, J.M.: Skilled independent control of individual motor units via a non-invasive neuromuscular–machine interface. Journal of

Neural Engineering **18**(6), 066019 (2021)

[33] Jiang, X., Dai, C., Liu, X., Fan, J.: Open access dataset and toolbox of high-density surface electromyogram recordings. PhysioNet (2021)

[34] Yang, S., Chen, C., Li, D., Zhu, X.: A consecutive multi-day high-density surface electromyography dataset comprising 7 grasps and 11 gestures. Scientific Data **12**(1), 1420 (2025)

[35] Hyvarinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE transactions on Neural Networks **10**(3), 626–634 (1999)

[36] Holobar, A., Zazula, D.: Multichannel blind source separation using convolution kernel compensation. IEEE Transactions on Signal Processing **55**(9), 4487–4496 (2007)

[37] Avrillon, S., Hug, F., Baker, S.N., Gibbs, C., Farina, D.: Tutorial on muedit: An open-source software for identifying and analysing the discharge timing of motor units from electromyographic signals. Journal of Electromyography and Kinesiology **77**, 102886 (2024)

[38] Murks, N., Škarabot, J., Kramberger, M., Divjak, M., Sedej, G., Valenčič, T., Connelly, C.D., Thomason, H., Holobar, A.: The efficiency of manual editing of high-density surface electromyogram decomposition depends on the recorded muscle and contraction level but less on the operator's experience. IEEE Transactions on Neural Systems and Rehabilitation Engineering (2025)

[39] Klotz, T., Rohlén, R.: Revisiting convolutive blind source separation for identifying spiking motor neuron activity: From theory to practice. arXiv preprint arXiv:2502.04065 (2025)

[40] Farina, D., Arendt-Nielsen, L., Merletti, R., Indino, B., Graven-Nielsen, T.: Selectivity of spatial filters for surface emg detection from the tibialis anterior muscle. IEEE transactions on biomedical engineering **50**(3), 354–364 (2003)

[41] Minetto, M.A., Holobar, A., Botter, A., Farina, D.: Discharge properties of motor units of the abductor hallucis muscle during cramp contractions. Journal of neurophysiology **102**(3), 1890–1901 (2009)

[42] Dideriksen, J.L., Negro, F., Enoka, R.M., Farina, D.: Motor unit recruitment strategies and muscle properties determine the influence of synaptic noise on force steadiness. Journal of neurophysiology **107**(12), 3357–3369 (2012)

[43] Lundsberg, J., Björkman, A., Malesevic, N., Antfolk, C.: Inferring position of motor units from high-density surface emg. Scientific Reports **14**(1), 3858 (2024)

[44] Buchthal, F., Erminio, F., Rosenfalck, P.: Motor unit territory in different human

muscles. Acta Physiologica Scandinavica **45**(1), 72–87 (1959)

[45] Semmler, J.G., Nordstrom, M.A., Wallace, C.J.: Relationship between motor unit short-term synchronization and common drive in human first dorsal interosseous muscle. Brain research **767**(2), 314–320 (1997)

[46] Mechler, F.: Matlab Code of Hartigans Dip Test [FORTRAN]. http://nicprice.net/diptest/

[47] Vázquez Noguera, J.L., López Pezoa, E., Espínola, G., Grillo, S.A.: A symmetric purity measure for clustering comparison. Annals of Data Science, 1–15 (2025)

[48] Jaccard, P.: Étude comparative de la distribution florale dans une portion des alpes et des jura. Bull Soc Vaudoise Sci Nat **37**, 547–579 (1901)

[49] Jaccard, P.: The distribution of the flora in the alpine zone. 1. New phytologist **11**(2), 37–50 (1912)

[50] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)

[51] BenSaïda, A.: Shapiro-Wilk and Shapiro-Francia Normality Tests. https://www.mathworks.com/matlabcentral/fileexchange/13964-shapiro-wilk-and-shapiro-francia-normality-tests

[52] Farina, D., Enoka, R.M.: Surface emg decomposition requires an appropriate validation. Journal of neurophysiology **105**(2), 981–982 (2011)

[53] Zhu, J., Feng, F., Ni, K., Zheng, Y., Zhang, J.: Four subtypes of motor neurons exhibiting mutually exclusive firing patterns in the spinal ventral horn. Neuroscience letters **525**(2), 163–167 (2012)

[54] ALIVE, M.: Description and analysis of the emg signal

[55] Del Vecchio, A., Negro, F., Holobar, A., Casolo, A., Folland, J.P., Felici, F., Farina, D.: You are as fast as your motor neurons: speed of recruitment and maximal discharge of motor neurons determine the maximal rate of force development in humans. The Journal of physiology **597**(9), 2445–2456 (2019)

[56] Meng, L., Hu, X.: Cross-person decomposition of surface electromyogram for efficient motor unit activity predictions. Journal of Neural Engineering **22**(4), 046052 (2025)

[57] Henneman, E.: Relation between size of neurons and their susceptibility to discharge. Science **126**(3287), 1345–1347 (1957)

[58] De Luca, C.J., Contessa, P.: Biomechanical benefits of the onion-skin motor unit control scheme. Journal of biomechanics **48**(2), 195–203 (2015)

[59] Fu, J., Zhang, S., Rakhshan, M., Helen, H.J., Wen, Y.: Consistency analyses of open-source software for motor unit decomposition using high-density electromyography. In: Proceedings of the IEEE EMBS Conference on Neural Engineering (NER) (2025)

[60] Weinman, L.E., Del Vecchio, A., Mazzo, M.R., Enoka, R.M.: Motor unit modes in the calf muscles during a submaximal isometric contraction are changed by brief stretches. The Journal of Physiology **602**(7), 1385–1404 (2024)

[61] Madarshahian, S., Letizi, J., Latash, M.L.: Synergic control of a single muscle: The example of flexor digitorum superficialis. The Journal of Physiology **599**(4), 1261–1279 (2021)

[62] Kirsch, E., Rivlis, G., Schieber, M.H.: Primary motor cortex neurons during individuated finger and wrist movements: correlation of spike firing rates with the motion of individual digits versus their principal components. Frontiers in neurology **5**, 70 (2014)

[63] Duchateau, J., Enoka, R.M.: Human motor unit recordings: origins and insight into the integrated motor system. Brain Research **1409**, 42–61 (2011) https://doi.org/10.1016/j.brainres.2011.06.011

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.tex