Pattern Matching for Real-Time Extraction of Fast and Slow Spectral Components From sEMG Signals

Alvaro Costa-Garcia, Akihiko Murai, and Shingo Shimoda, *Member, IEEE*

***Abstract*— Previous studies have demonstrated the potential of surface electromyography (sEMG) spectral decomposition in evaluating muscle performance, motor learning, and early diagnosis of muscle conditions. However, decomposition techniques require large data sets and are computationally demanding, making their implementation in real-life scenarios challenging. Based on the hypothesis that spectral components will present low inter-subject variability, the present paper proposes the foundational principles for developing a real-time system for their extraction by utilizing a pre-defined library of components derived from an extensive data set to match new measurements. The model library was tailored to fulfill specific requirements for real-time system application and the challenges encountered during implementation are discussed in the paper. For system validation, four distinct data sets comprising isotonic and isometric muscle activations were utilized. The extracted during validation showed low inter-subject variability, suggesting that a wide range of physiological variations can be described with them. The adoption of the proposed system for muscle analysis could provide a deeper understanding of the underlying mechanisms governing different motor conditions and neuromuscular disorders, as it allows for the measurement of these components in various daily-life scenarios.**

***Index Terms*— Electromyography, spectral component, motor control, muscle activity, muscle fatigue, real-time systems.**

I. Introduction

SURFACE Electromyography (sEMG) records electrical signals on the skin that represent the summation of electrical potentials produced by active muscle fibers contributing to muscle contraction. The spectral properties of these potentials are affected by various neural, physiological, and experimental factors, such as the firing rate of motor neurons [1], [2], the number and size of recruited motor units, the physical and metabolic properties of active muscle fibers [3], [4], the volume conduction properties of muscle, fat, bone, and skin layers [5], and the arrangement of recording electrodes [6].

Researchers have proposed evaluating the spectral properties of sEMG signals to infer internal muscle properties, such as motor unit size and muscle fiber type distributions [7], [8], throughout history. However, this remains a highly debated topic due to the numerous reported parameters associated with changes in sEMG signal frequency domains. An interesting summary of this debate can be found in the point and counterpoint discussion between Farina [9].

In 2004, Wakeling and Rozitis found that the spectrum of raw sEMG signals recorded during short isometric contractions over the leg extensor muscle could be largely described by two principal components in the spectral domain [10]. In 2006, Tscharner and Goepfert extracted very similar components from the gastrocnemius and tibialis anterior muscles on isotonic sEMG data recorded during running [11]. In both works, the terms "fast" and "slow" were used to describe the significant differences in the median frequency of each of the extracted components. The authors hypothesized that these components might represent families of motor unit action potentials (MUAPs) with significantly different properties, as the spectral distribution of both components corresponded to that reported from these potentials [12].

In 2022, to further test the nature of these components, Costa-García et al. found that multi-channel sEMG data recorded from the biceps muscle during long and high-force fatiguing isometric contractions can also be described by a fast and slow spectral component with a >80% reconstruction rate [13]. Both component showed inter-subject and inter-electrode stability. Moreover, the analysis performed in this paper showed that both components exhibit a specific behavior during fatigue; the fast component dominated the beginning of the contraction, but its contribution decreased with fatigue, while it was progressively replaced by the slow frequency component. The same behavior was reported for nine healthy subjects, demonstrating the potential benefits of this component analysis for the quantification of muscle fatigue.

In the previous studies the extraction of these components (see Figure 1 for the conceptual decomposition flow) has mostly been performed offline, requiring large data sets to cover the diversity of scenarios encountered in daily life. Moreover, the computational demands of analyzing these

Manuscript received 7 June 2023; revised 31 July 2023; accepted 29 August 2023. Date of publication 1 September 2023; date of current version 12 September 2023. This work was supported in part by the Toyota Motor Corporation and in part by the Kakenhi National Research Grants. *(Corresponding author: Alvaro Costa-Garcia.)*

Alvaro Costa-Garcia and Akihiko Murai are with the Co-Creative Platform Research Team, Human Augmentation Research Center, AIST Kashiwa II Campus, The University of Tokyo, Kashiwa, Chiba 277-0882, Japan (e-mail: alvaro.costagarcia@aist.go.jp).

Shingo Shimoda is with the Graduate School of Medicine, Nagoya University, Nagoya, Aichi 466-8550, Japan (e-mail: sshimoda@ieee.org).

Digital Object Identifier 10.1109/TNSRE.2023.3311037

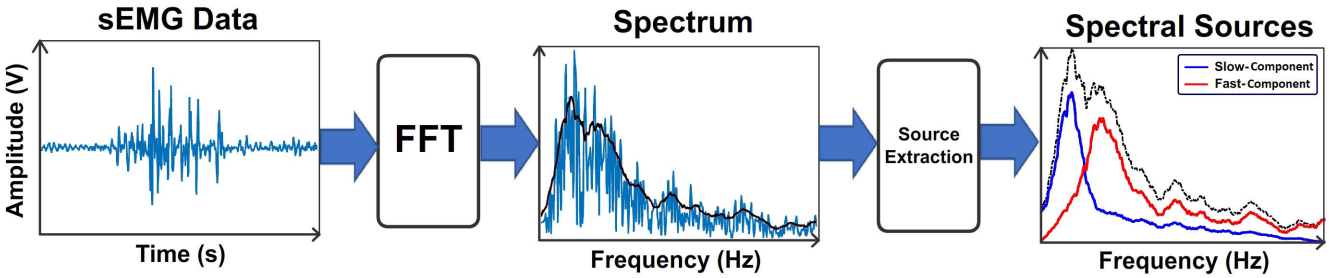

Fig. 1. sEMG signal decomposition. Conceptual decomposition of the envelope of the spectrum of an sEMG signal into fast and slow spectral components.

data sets difficult their online computation. Reducing these obstacles will enable the measurement of such components in meaningful situations like the monitoring the age-related decay of muscle mass known as sarcopenia [14] and other progressive motor conditions [15], [16], [17] which early diagnosis is critical for their treatment [18], [19], [20].

In this paper, we aim to develop an online system capable of extracting these components in real-time, using a pre-defined library of patterns obtained from a large data set. Our hypothesis is that, given the inter-subject and inter-electrode component stability infered by the comparison of previous work results, a comprehensive library of patterns can be generated to cover most scenarios encountered in daily life, and new measurements can be matched to the closest pattern in the library, thus allowing the real-time extraction of component contributions. By developing such system, our research aims to bridge the gap between theoretical understanding and practical applications. The online extraction of the fast and slow components could offer several significant benefits, including real-time estimation of biological states such as muscle fatigue and motion intentions. These factors are crucial for monitoring appropriate training levels for rehabilitation from motion paralysis and supporting movement with robots. We further suggest that such a system could provide valuable insight into the origin of these components and their physiological significance, as it would enable their measurement in a variety of daily-life scenarios, potentially leading to a better understanding of the mechanisms underlying different neuromuscular disorders.

The next section will detail the materials and methods employed to define and validate the proposed real-time system. Initially, a straightforward conceptual design will be presented (Figure 2), followed by the identification and resolution of challenges that arose during implementation. Subsequently, the solutions used to address each challenge will be validated by analyzing four distinct data sets that include sEMG signals recorded from isometric and isotonic contractions.

II. Materials and Methods

A. Basic Design of the Real-Time System

Figure 2 depicts the overall workflow of the real-time system designed for the purpose of this study. The system assumes that any set of sEMG spectra can be decomposed into a linear combination of two normalized spectral components. Once these components are identified, they can be utilized as a model to extract the weights that modulate a

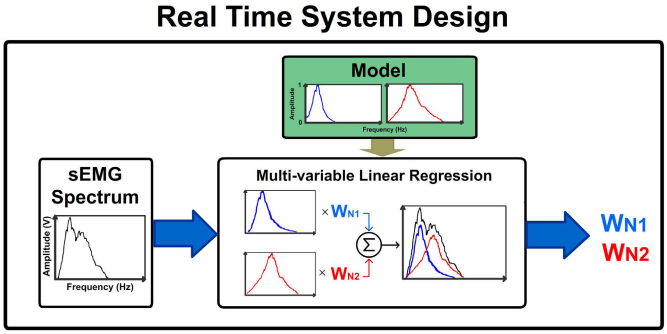

Fig. 2. Real-Time System Design. The system is designed to extract the relative contribution of slow and fast frequencies from an sEMG signal by fitting a pre-defined model composed of two spectral components into the envelope of the signal's spectrum using multivariable linear regression.

new incoming sEMG spectrum. However, the methodology for component extraction developed in previous works [10], [11], [13] was highly task-specific, and its generalization in the form described in Figure 2 is challenging. Throughout this paper, the authors will elaborate on and address all these challenges in defining a functional system. Three groups of subjects were recorded with sEMG signals from biceps and gastrocnemius muscles, including both isometric and isotonic contractions. This data was then organized into four data sets used to validate each stage in the definition of the proposed system.

B. Data Sets, Participants and Experiments

1) Isometric Data: For the isometric data, an elastic medium-density electrode band was utilized to record sEMG signals around the forearm (as shown in figure 3A). The band, developed by Oisaka electronic equipment ltd [21], consisted of 25 dry active electrodes arranged in five arrays with an inter-array distance of 3.75 cm. The band was grounded via a wristband. EMG signals were acquired as the differential signal between each pair of vertically consecutive electrodes, with an inter-electrode distance of 2 cm, resulting in a 5 × 4 matrix of spatially distributed signals. Only the 3 middle arrays (electrodes 5-16) were selected for recording as they include the bipolar channels withing the motor point area of the bicep-brachii muscle according to the guidelines of the Surface Electromyography for the Non-Invasive Assessment of Muscles (SENIAM) project [22]. The signals were digitized at a sampling rate of 2000 Hz and transmitted to a PC through a USB connector. The electronic components were powered by

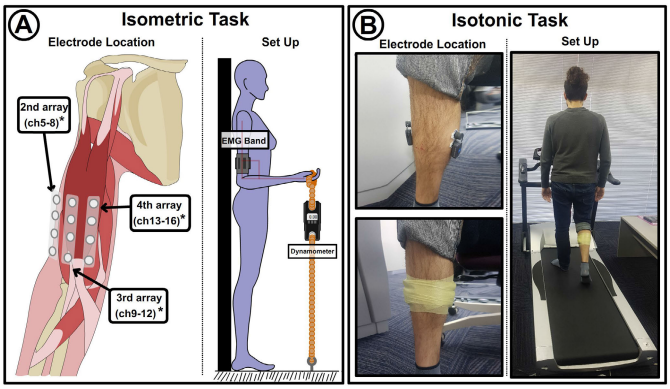

Fig. 3. Experiment Setups. **a)** Isometric contractions of the biceps muscle were recorded using a 3 × 4 matrix of bipolar electrodes. Participants stood against a wall with their elbow flexed at 90° and pulled on a string attached to a dynamometer. **b)** Isotonic data from the gastrocnemius muscle was recorded using a bipolar electrode during walking on a treadmill at 2 km/h.

a small battery and did not include hardware filters. The use this channel distribution comes from a previous study where the effects of spatial distribution on the spectral patterns was tested [13]. Participants were instructed to sustain an isometric contraction of their biceps at 70% of their maximum force (previously measured) for 30 seconds while receiving real-time feedback from a dynamometer located between the subject and the floor. Each participant completed this task four times, resulting in data set 1. The experiment included nine healthy, right-handed individuals: 4 women and 5 men aged 27 to 45 years (34.88 ± 6.71).

2) Isotonic Data: For the isotonic data, sEMG recordings were obtained from a single bipolar electrode placed on the surface of the lateral gastrocnemius muscle following the guidelines of the Surface Electromyography for the Non-Invasive Assessment of Muscles (SENIAM) project [22]. Participants were asked to walk on a treadmill at 2 km/h for 10 minutes, and each participant completed 4 trials. Two groups of subjects participated in the experiment.

The first group consisted of seven healthy individuals (4 women and 3 men) with ages ranging from 29 to 48 years (39.29 ± 7.57). Each participant completed a session of the experiment on four different days, divided into two groups. Two sessions used a thin elastic band to reinforce the electrode to the leg (as shown in figure 3B) to reduce motion-induced vibration of the electrode (data set 3), while the other two sessions were done without the use of the band (data set 2). This was done to evaluate the possible appearance of motion artifacts on the sEMG data.

The second group consisted of nine healthy individuals (5 women and 4 men) with ages ranging from 28 to 48 years (38.67 ± 7.91). Each participant completed a single session of the experiment using the elastic band to reinforce the electrodes, resulting in data set 4.

The number of participants in each data set was chosen to align with those utilized in the aforementioned studies for spectral component extraction [11], [13]. In total, nine subjects participated in the isometric recording, and sixteen subjects were involved in the isotonic experiments, with only four individuals taking part in both types of experiments. Selection criteria required all participants to have not known motor conditions and healthy life-styles that involve at least two hours of weekly exercise. Prior to participation, all subjects were provided with comprehensive information regarding the experimental procedures, and they provided informed consent following the guidelines outlined in the Declaration of Helsinki.

3) Data Sets Usage: Data sets 1 and 2 will be utilized to assess the differences in the targeted spectral distribution of sEMG signals under different conditions: data set 1 corresponds to isometric data with a low artifact environment, while data set 2 represents isotonic data without the use of a reinforcement band, where artifacts may influence the recorded data. Based on this analysis, signal filtering for data sets 3 and 4 (isotonic with reinforcement band) will be determined. Subsequently, data set 3 will serve as the training data for generating the model library, while data set 4 will be used for validation purposes.

C. System Comparison

1) System From Previous Work: Figure 4A shows the conceptual design of the system developed in Costa-García et al. [13] which will be used as reference for current work. During isometric contractions, a set of sEMG segments was recorded and subjected to a Non-negative Matrix Factorization, which resulted in the extraction of fast and slow spectral components. This factorization enabled the decomposition of n spectral sources $En(f)$ using the formula:

$$En(f) = wn1 \cdot H1(f) + wn2 \cdot H2(f) \tag{1}$$

where $H1(f)$ and $H2(f)$ are the normalized components, and $wn1$ and $wn2$ are the weight indicating the degree to which each component contributed to the spectrum $En(f)$.

2) Real-Time System: According to the real-time system described in Figure 2, modulation weights are extracted from an sEMG spectrum using multivariable linear regression after fixing a set of normalized components, $H1(f)$ and $H2(f)$. In this paper, the term "model" refers to this set of two components. Figure 4B illustrates the conceptual design of the system proposed in this study, which is divided into three stages. The first stage involves the use of a large amount of data from different subjects to extract a model library composed of a set of normalized spectral components, $H1(x, f)$ and $H2(x, f)$. This library of x models assumes that the shape of both components will show certain inter-subject variability. In the second stage, a reduced set of sEMG segments is used as training data for system calibration, which involves selecting the single model from the library that best represents the new data. Finally, in the third stage, the selected model is applied to new incoming data to obtain the weights, $wn1$ and $wn2$, by performing multivariable linear regression [23].

3) Implementation Challenges: In order to transition from the system presented in Costa et al. (Figure 4A) to the proposed system (Figure 4B), several modifications are necessary.

The first challenge is to generalize the data under analysis and account for the differences between isometric and isotonic contractions. The proposed system requires sEMG data recorded from active muscles. In the work of Costa et al., the data under analysis consisted of isometric contractions, where the muscles are continuously contracted. However, isotonic motions involve periods of active and inactive muscle

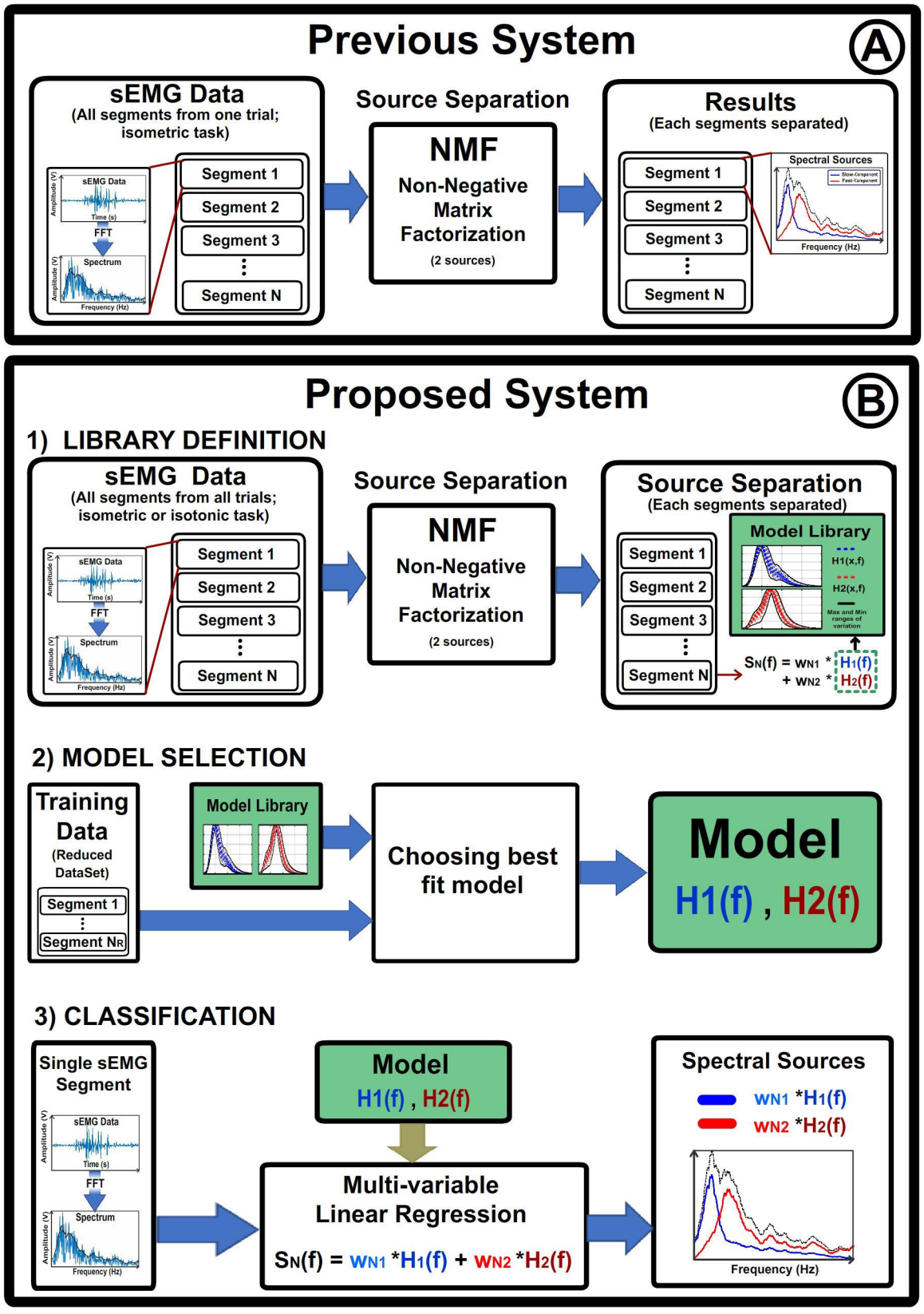

Fig. 4. Comparison of systems. **a)** System introduced in [13] involves extracting epochs from sEMG data recorded during a single session. These epochs then undergo a Non-negative Matrix Factorization algorithm in order to extract both slow and fast spectral components. **b)** New system allows for real-time extraction of component contribution. In the first stage, a model library is defined based on a large data set by applying Non-negative Matrix Factorization to different groups of segments within the data. In the second stage, a reduced set of training data is used to extract the components that best represent the new data from the model library. Finally, the extracted model is used for real-time classification of sEMG using Multi-variable Linear Regression.

contraction. Therefore, before extracting a model, a smart segmentation approach is required to select the data epochs that best represent the periods of muscle activity. Moreover, data acquired during periodic isotonic contractions may be affected by motion-induced artifacts that were not present in isometric data, where muscle contraction did not produce significant motion.

The second challenge is to define a model library. According to the results obtained by Costa et al., the normalized components $H1(f)$ and $H2(f)$ (recorded from the same muscle) did not show significant variability between subjects, which allows for the definition of a reduced library of models that can be applied to any subject. To create efficient models, it is necessary to strike a balance between their stability and how effectively they can reconstruct the original data.

The last challenge is to define a protocol to select the model from the library that best represents the new data that needs to be analyzed. As mentioned before, variations in the model between subjects, even small, still exist. This variability is manifested as different instances of the components contained in the model library. Therefore, by knowing the ranges of this variability, it will be possible to obtain from the library the instance of the model that best fits the new data. The methodology used to address each of these challenges is described in the following three sections.

D. First Challenge: Isometric VS Isotonic Data

1) Segmentation: The segmentation process differed depending on the type of data being analyzed. For sEMG signals recorded during isometric contractions (data set 1), 1-second epochs with 0.5 seconds of overlap between segments were used (30-second trials, see II-B). Each of the 12 channels was divided into 60 epochs, resulting in a total of $12 \times 60 = 720$ epochs per trial.

For isotonic sEMG signals (data sets 2-4), individual muscle activations were segmented using the methodology for smart periodic sEMG data segmentation introduced by Costa-García et al. in [24]. Although walking time and speed were fixed for this experiment, the total number of epochs varied between subjects (456 ± 34 epochs for each 10-minute walking trial, see II-B).

2) Noise Evaluation: Prior to extracting the sEMG spectral components $H1(f)$ and $H2(f)$, it was necessary to filter motion-induced artifacts coupled to the signal spectrum. To evaluate the spectral ranges affected by these artifacts, data sets 1, 2, and 3 were compared. Data set 1 consisted of epochs extracted from bicep isometric contractions where the electrodes did not experience significant movement. Data sets 2 and 3 contained epochs recorded from the lateral gastrocnemius during gait, with the difference that Data set 3 was recorded while subjects were using an extra elastic band to reinforce electrode attachment and reduce motion-induced noise (see Figure 3B). The absolute value of the Fast Fourier Transform was computed for all epochs in each data set, and their average values were compared to determine the frequencies affected by motion-induced artifacts.

Based on this information, the filter design GUI from MATLAB was used to define an Infinite Impulse Response (IIR) Butterworth band-pass filter for removing motion noise during the pre-processing stage.

E. Second Challenge: Definition of the Model Library

The spectral components $H1(f)$ and $H2(f)$ that form our model were extracted from the decomposition of the spectral shape of a set of segments into two sources by applying a Non-Negative Matrix Factorization [25].

1) Non-Negative Matrix Factorization: Prior to source separation, a moving average filter was applied to each spectrum to emphasize the contributions of the sources contained in their envelope. Subsequently, the spectrum of sEMG signals was described as:

$$E = W \cdot H \tag{2}$$

$$E = m \times f;\ W = m \times s;\ H = s \times t \tag{3}$$

where E is an $m \times f$ matrix of sEMG spectral data (with m the number of segments and f the frequencies under analysis), W is an $m \times s$ matrix containing the weights associated with each component used to reduce the m segments to an s-dimensional space, and H is an $s \times f$ matrix containing the spectral patterns ($H1(f)$ and $H2(f)$ in the case of this work). Matrices W and H can be calculated from E through the Non-negative Matrix Factorization (NMF) algorithm by fixing the s-dimensionality ($s = 2$, number of components). The NMF algorithm extracts the spectral components of H by minimizing the correlation between them [26].

2) Number of Segments: Once the number of components is decided ($s = 2$), the other parameter that needs to be fixed is the minimum number of segments needed for the extraction of valid spectral components (m number of segments according to equations 2-3). In this work, the validity of the spectral components $H1(f)$ and $H2(f)$ was evaluated by two factors. The first one was the Value Account For (VAF), which represents with a number from 0 to 100 how accurately $H1(f)$ and $H2(f)$ can be used to reconstruct the original data. As can be expected, the reconstruction rate will be higher for a reduced number of segments, and its value will drop for higher numbers. Values of VAF over 80% after the application of Non-Negative Matrix Factorization are generally accepted as high reconstruction rate [27]. The second parameter was the cross-correlation between the spectral patterns $H1(f)$ and $H2(f)$ extracted from different groups of segments. This parameter provides a number between -1 and 1. Value over 0.9 are associated to significantly similar signals [28] and will be used to represent the stability of the model. In this case, it is expected that the stability will be smaller for a small number of segments and larger for a higher m value. The balancing between the reconstruction accuracy and the stability of the model was used to choose the number of segments for the computation of the model.

sEMG signals from data set 3 were used to find the optimal number of segments following the process described in Figure 5. Temporal data was filtered to remove motion-induced noise using the information obtained from the noise evaluation (See II-D.2 and III-A). After that, individual muscle activations were segmented, and their spectra were computed as the envelope of the absolute value of the Fast-Fourier Transform.

Data set 3 was divided into groups of different m-number of segments, and different sets of spectral components $H1m(f)$ and $H2m(f)$ were computed from each group. The total

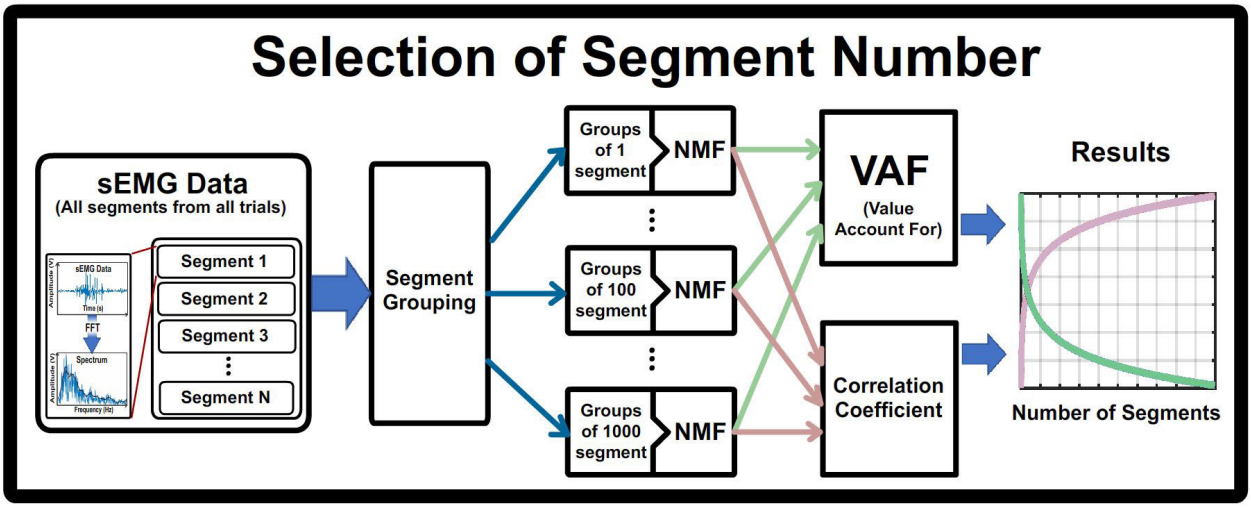

Fig. 5. Number of segments for Model Library definition. The total number of sEMG segments is divided into groups, ranging from 1 to 1000 segments per group. For each group, the Non-negative Matrix Factorization algorithm is applied to extract the model associated with the segments under evaluation. Finally, the Variance Accounted For (VAF) and the correlation coefficient extracted from each group are computed as indices representing the reconstruction accuracy and stability of the models obtained from each number of segments.

VAF associated with an m number of segments was computed as the average of the VAFs of each group. The stability of the model for the same number of segments was computed as the average of the cross-correlation coefficients obtained between each set of spectral components $H1m(f)$ and $H2m(f)$. Both values were computed for $m = [5, 10, 20, 50, 100, 200, 300, 500, 700, 1000]$ and graphically represented.

3) Model Library: Once the minimum number m of segments was fixed, the data from data set 3 was divided into x groups of m segments (as shown in Figure 6A), where $x = T/m$, T being the total number of segments, and m being the chosen number of segments for model extraction. After this, the NMF algorithm was used to extract a set of $H1(f)$ and $H2(f)$ spectral components, i.e., a model from each group (model library). The variations observed between models are a consequence of the natural physiological variabilities between humans, namely muscle fiber type distribution rates, volume conduction on skin and fat layers, inter-subject motor strategy changes, etc. The accurate characterization of all these sources is complex and time-consuming for a real-time system that aims for short and feasible training times. Therefore, the current work opted for defining a library of models whose variations are expected to be within a wide physiological range.

4) Validation of Model Library: To validate this library, a 7-fold cross-validation was conducted on both components (Figure 6B). Since the greatest variations between components are expected across data from different subjects, each fold was obtained by separating the $Hy(f)$ components from one subject from the $Hz(f)$ components of the remaining subjects, where $y + z = x$ and x represents the total number of models contained in the library (seven folds for seven subjects). Each $Hy(f)$ component was compared with all $Hz(f)$ components, and the maximum VAF was chosen to represent how well $Hy(f)$ can be reconstructed by components extracted from other subjects. The average and standard deviation of the VAF values (VAFy) were calculated as an index showing the reconstruction potential for each fold. The average of these indexes was then computed from the seven folds for both spectral components. Figure 6B illustrates an example of this process applied to the low-frequency component $H1(f)$. This process was applied using only folds extracted from data set 3 (the set used to define the model library). The goal was to test the internal coherence between the spectral components extracted from different subjects.

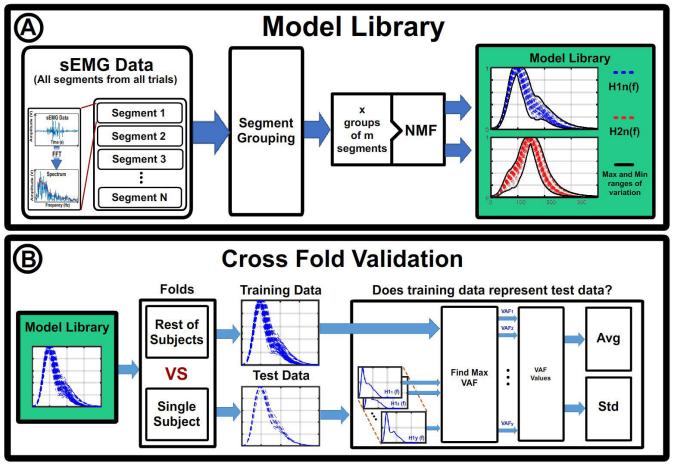

Fig. 6. Model Library and Validation. **(a)** Computation of the Model Library. The data segments were divided into *x* groups of *m* segments, and a Non-negative Matrix Factorization algorithm was used to extract a model from each group. The group of *x* models extracted constitutes the Model Library. **(b)** Validation of the Model Library. The Variance Accounted For (VAF) was used to evaluate how well the models extracted from each subject can be reconstructed by models from the rest of the subjects. A cross-fold validation using the same number of folds as the number of subjects included in the original data was used to compute the average and standard deviation of the reconstruction index.

F. Third Challenge: Selecting the Best Fit Model

1) Selection Algorithm: Figure 7a illustrates the process used to select the best fit model from the library. First, a new segment of validation data (from data set 4) undergoes the

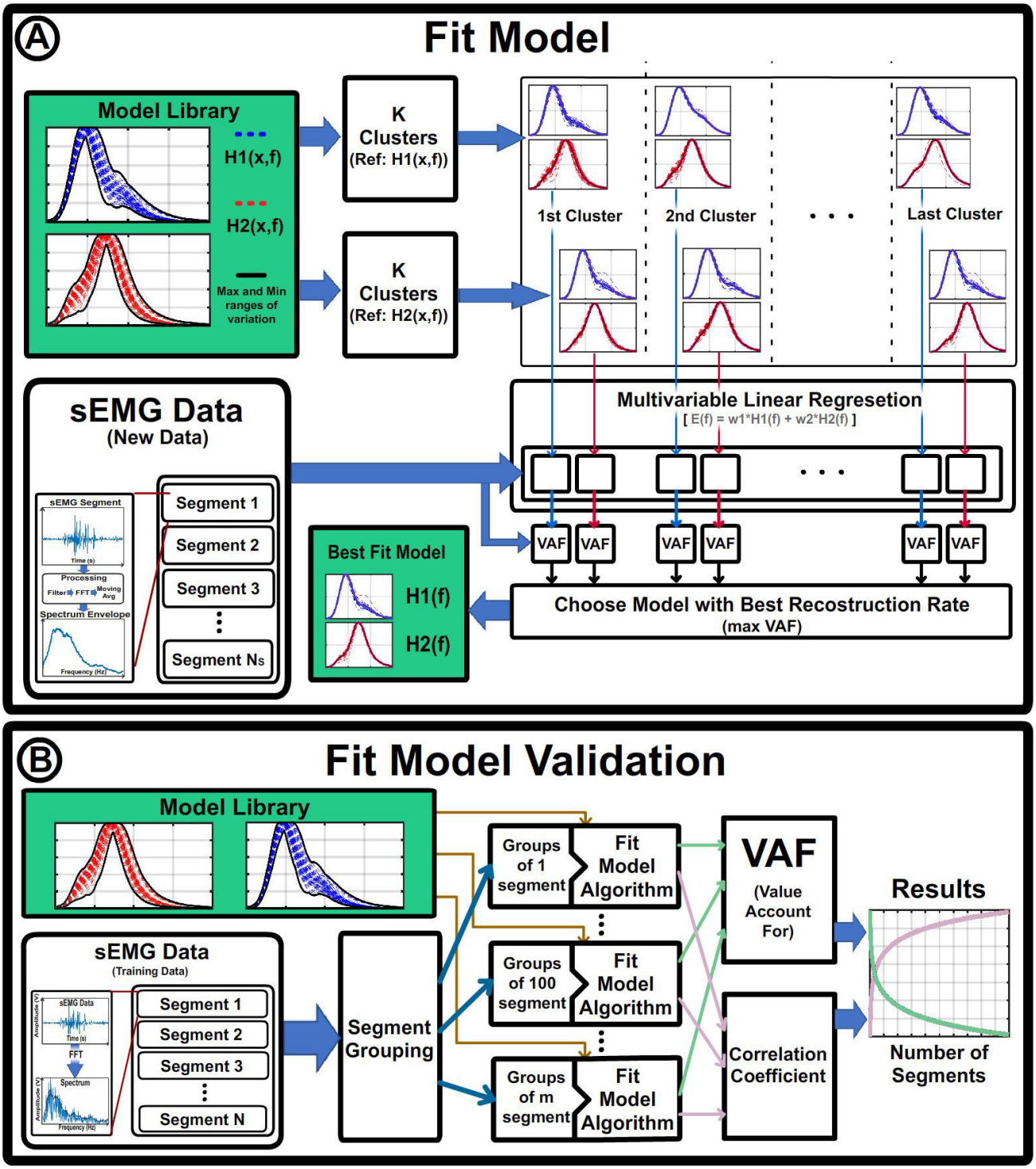

Fig. 7. Fit Model Algorithm and Validation. **a)** Algorithm. The *H*1 and *H*2 components included in the Model Library are k-clustered and averaged. Each averaged cluster is used to reconstruct a set of new data using Non-negative Matrix Factorization. The averaged cluster providing the highest VAF is selected as the best fit model. **b)** Algorithm Validation. The algorithm is used to extract a model from different numbers of sEMG data segments in a process similar to the one used to select the number of segments for the Model Library. Finally, the reconstruction accuracy and stability are evaluated using VAF and correlation coefficient.

same pre-processing as the data used to obtain the model library (data set 3), which includes noise removal bandpass filtering, spectral computation using the absolute value of the Fast-Fourier Transform, and envelope extraction by a moving average filter. Next, the $H1(x, f)$ and $H2(x, f)$ components contained in the library are divided into 10 different groups based on their similarity. To accomplish this, a k-means clustering algorithm with $k = 5$ is used to divide the matrix of $H1(x, f)$ spectral components into 5 clusters (with their respective counterparts of $H2(x, f)$). The same process is repeated using $H2(x, f)$ components as reference for the clustering algorithm, obtaining another 5 clusters. The average values of the 10 clusters are considered as candidate models for the new incoming data. Then, a multivariable linear

regression is applied to fit the new data to Equation 1, where $E(f)$ represents each reconstructed segment of process data, and $H1(f)$ and $H2(f)$ are each averaged cluster. Finally, the reconstruction rate (VAF) is computed by comparing the original and reconstructed data. The set of $H1(f)$ and $H2(f)$ providing the highest VAF is selected as the best fit model. Depending on the amount of data contained in the library and the observed variability of $H1(f)$ and $H2(f)$ components, the number of clusters k can be modified for a more accurate extraction of the model. After the model is obtained from the new data, it can be used to evaluate the new incoming data from the same muscle over a range of motor tasks.

2) Algorithm Validation: Section II-E.2 defined a methodology to choose the minimum number of segments m required for the computation of the models forming the model library so that they fit certain reconstruction accuracy and stability constraints. Using that library, the method described in Figure 7a allows the extraction of a model from any ns number of segments from a set of new data. For this method to be a useful tool for the definition of a model from a reduced data set, the number of segments of the new data (ns) must be smaller than the number of segments used for the definition of the library models (m) without significantly affecting the reconstruction accuracy and stability of the model.

Data set 4 was used as validation data using the methodology described in Figure 7b. sEMG segments were divided into different numbers of groups of ns segments. For each group, the extraction of the model was done using the process described in Figure 7a. Model reconstruction accuracy was computed as the VAF by comparing the original data and the data reconstructed from the best fit model after applying a multivariable linear regression, and model stability was obtained from the average cross-correlation between the models obtained from different groups of ns segments. This process was repeated for different $ns < m$ numbers of segments to show how much the number ns can be reduced without compromising stability and reconstruction accuracy.

III. Results

A. Noise Removal

Figure 8a shows the average spectrum computed from data sets 1-3. Signals recorded during bicep isometric contraction (data set 1) presented significantly higher amplitude than those recorded on gastrocnemius muscle during gait (data sets 2-3). This was due to the inherent differences in the force requirements of each type of motion. Therefore, for proper visualization of the frequency bands affected by noise, the averaged spectra from each set were normalized based on their activity at 50 Hz, which represented the most active frequency during isometric contraction. According to Figure 8a, motion-induced noise strongly affects frequencies under 35 Hz, with this effect being lower in data set 3, where the sEMG electrode position was reinforced by an elastic band (Figure 3b). High frequencies also showed an increased activation on gait data after 70 Hz. However, this activity does not present significant differences between data sets 2 and 3, implying that the reinforcement of electrodes does not contribute to the reduction of the noise in this band. Based on these results, an Infinite Impulse Response (IIR) Butterworth bandpass filter was designed to maintain the sEMG frequencies between 35 Hz and 70 Hz. Figure 8b shows a representation of its magnitude response in decibels (dB). Figure 8c shows the comparison between the artifact-free isometric data (data set 1, black-dotted line) with the result of applying the designed filter to data set 3 (green line). In order to reduce the effect of artifacts and a consistent relation between data sets, both data set 3 (used to define the model library) and data set (used for validation) underwent the same filtering stage.

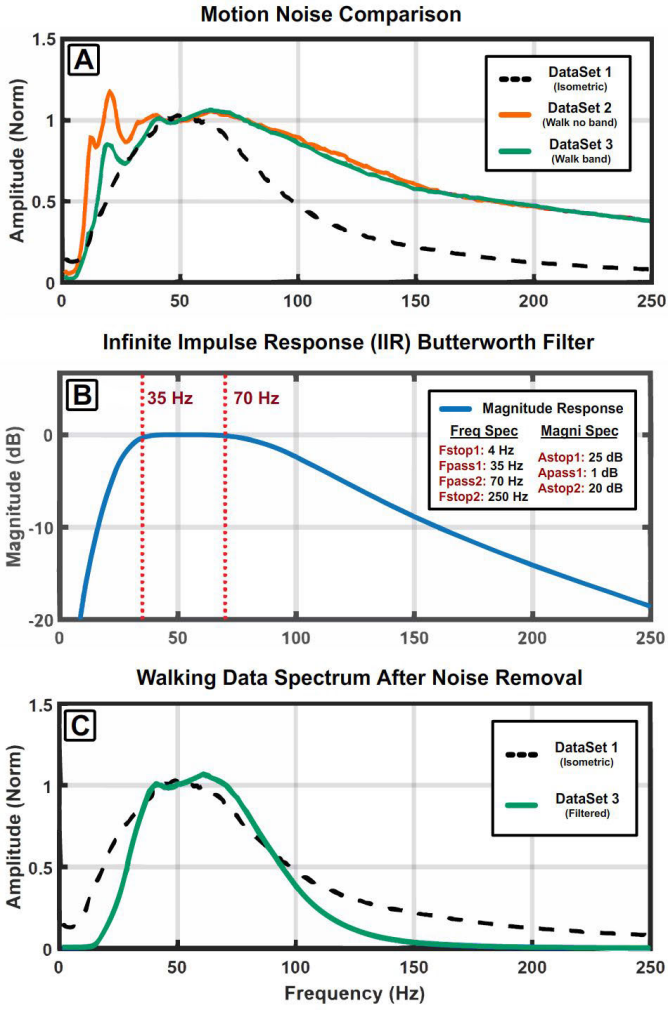

Fig. 8. Noise Filtering. **(a)** Average sEMG spectrum from raw data extracted from data sets 1 (black dotted), 2 (orange) and 3 (green). **(b)** Bandpass filter between 35 and 70 Hz designed to remove the spectral range with higher motion noise contributions. **(c)** Average sEMG spectra from data sets 1 (black dotted) and 3 (green) after the application of the noise removal filter.

B. Number of Segments for Model Library

Figure 9 shows the averaged values of the correlation coefficient and VAF extracted from groups of spectral components $H1(f)$ and $H2(f)$ computed from different number of data segments (from data set 3). Both the model reconstruction accuracy (given by VAF) and its stability (given by the cross-correlation coefficient) present exponential changes for small number of segments. This behavior becomes almost linear after 100 segments. Therefore, $m = 300$ was selected for the definition of the model library as it was contained within the range of linear variation of the parameters evaluated and also presented acceptable balance between the reconstructions

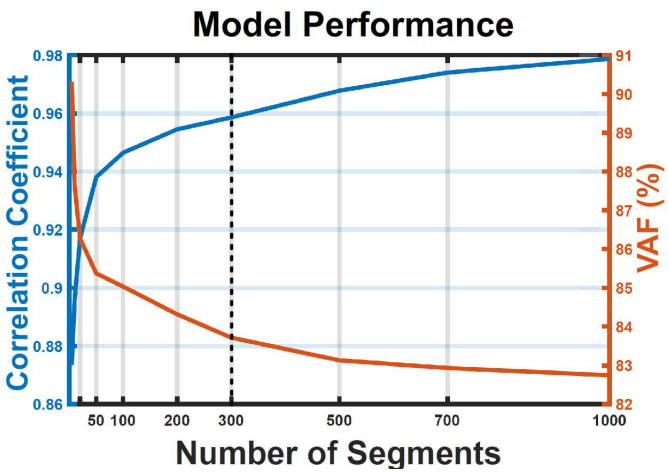

Fig. 9. Model Performance. Estimation of model stability and reconstruction accuracy quantified by correlation coefficient (blue line) and the VAF (orange line) for a different number of segments. The vertical back-dotted lines represent the number of segments chosen for model extraction.

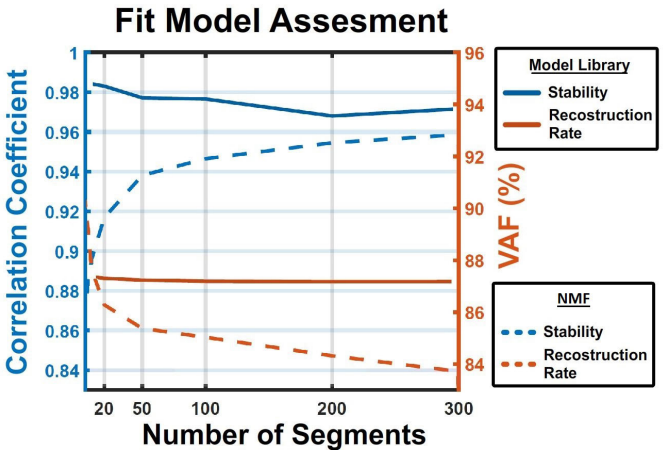

Fig. 10. Assessment of Model Fitting. Comparison between stability (blue lines) and reconstruction accuracy (orange line) of the model extracted using both the Non-negative Matrix Factorization (dotted lines) and the model extracted using the pre-computed Model Library and the selection algorithm presented in Figure 7 (straight lines).

accuracy (83.71%) [27] and stability (0.96) [28] of the model extracted from the NMF algorithm. Moreover, based on this $m = 300$ number, the $ns > m$ number of segments introduced in II-F.2 for the validation of the best fit model was tested within the range $ns = [10,\ 20,\ 50,\ 100,\ 200,\ 300]$.

C. Model Library Validation

Before testing the accuracy of the model to classify the validation data from data set 4. The model library obtained from data set 3 from groups of $m = 300$ segments was further tested by the cross-fold validation described in Figure 6B. Results show that components extracted from a single subjects could be described by the components extracted from the rest of them with reconstruction rates of $95.56 \pm 0.34\%$ in the case of $H1(f)$ and $95.42 \pm 1.87\%$ in the case of $H2(f)$. This results strengthen inter-subject consistency between the fast and slow spectral components.

D. Number of Segments for Choosing Best Fit Model

Figure 10 shows the reconstruction rate (VAF, Right Y-axis) and component stability (cross-correlation coefficient, Left Y-axis) from the spectral components extracted from the validation data (data set 4). X-axis shows the number of segments used for component extraction. Dotted lines represent the results of component extraction based on the direct application of the NMF to data set 4, while straight lines show the results of component extraction based on the model library defined from data set 3 (also applied to data set 4). This results show that the use of the library allows the extraction of spectral components from a reduce set of new data (any number between 20 and 300 segments) with higher stability and reconstruction rate, than the one obtained by the only use of the NMF algorithm. This can be expected when applying the NMF to a reduce number of segments as the redundancy between them might be not high enough for the methodology to extract common sources. In our results, this redundancy was already comprised from data set 3 into the model library and therefore, the extraction of components from the model library results in higher reconstruction rates and stability.

IV. Discussion

This study demonstrated the possibility of extracting fast and slow components from sEMG signals without requiring a direct application of source separation techniques. Source separation was only utilized once for defining the model library; subsequently, components can be extracted from the library using a simple multivariable linear regression.

The inherent differences between isotonic and isometric muscle contractions required motion-specific preprocessing stages for the proper extraction of components. The goal of this preprocessing was to ensure that the data under analysis contained the maximum possible information about active muscle fibers. In the temporal domain, this means segmenting epochs containing full periods of muscle contraction. In the spectral domain, the removal of motion-induced artifacts observed under 35 Hz and over 70 Hz (Figure 8A) was necessary on isotonic data to ensure that the spectral components extracted by the NMF algorithm fall within the spectral range dominated by active muscle fibers (Figure 8C) [29].

The model library was carefully designed to fulfill a set of requirements for its real-time system application. The models contained in the library were extracted from sets of 300 data segments that showed a balanced performance between sEMG reconstruction accuracy and component stability (Figure 9). In addition, the model components showed low inter-subject variability, suggesting that data from a reduced number of subjects can describe a wide range of physiological variations within the extracted components. Finally, the use of a previously defined model library allows for the extraction of spectral components from a reduced set of new data without compromising the model's stability and reconstruction rate (Figure 10).

The results presented in this paper provide a robust framework for the real-time extraction of the contribution of fast and slow spectral components to sEMG signals. The development of a model library and the implementation of motion-specific preprocessing stages are key factors to ensure accurate and stable model extraction. The use of a reduced set data for model extraction, made possible by the use of the model library, does not compromise the performance of the model

in terms of stability and reconstruction rate. In addition, the results suggest that spectral components can be extracted from only 20 muscle activations, enabling the application of this framework to a wide range of daily motions. Once component are extracted, they can be used for real-time classification with the application of a simple preprocessing stage and a multivariable linear regression.

Only data from the isotonic contraction of the gastrocnemius muscle during walking was used for the testing stage of the described methodology. To further validate the system, it will be necessary to expand the scope of data to include other muscles and contractions. Additionally, considering that wavelets have demonstrated better physiological description of sEMG signals [30], the system's performance should be re-tested after replacing Fast-Fourier Transformation with Wavelet analysis in the spectral computation stage. In future studies, this system will be utilized for measuring muscle fatigue in various daily situations. Furthermore, the authors are planning to employ this real-time approach for daily monitoring of muscle conditions in patients with sarcopenia. In sarcopenia, the progressive decrease of muscle mass has been linked to the decline of fast-twitch muscle fibers, a phenomenon similar to that observed during fatiguing contractions. Consequently, we expect to identify comparable long-term changes in the spectral components of sarcopenia patients, allowing for early diagnosis of this disease.

Our research strives to narrow the gap between theoretical comprehension and practical implementation by developing an online system that allows for real-time extraction of both fast and slow components from sEMG signals. The direct benefits of the presented system are twofold: firstly, it enables the continuous monitoring of muscle fatigue and motion intentions in real-world scenarios, allowing for timely adjustments in training levels and potentially enhancing rehabilitation strategies for motion paralysis. Secondly, it provides a valuable tool for investigating the origin and physiological significance of the fast and slow components, paving the way for advancements in the diagnosis and treatment of various neuromuscular disorders.

REFERENCES

[1] B. K. Barry, M. A. Pascoe, M. Jesunathadas, and R. M. Enoka, "Rate coding is compressed but variability is unaltered for motor units in a hand muscle of old adults," *J. Neurophysiol.*, vol. 97, no. 5, pp. 3206–3218, May 2007.

[2] E. Stalberg, "Propagation velocity in human muscle fibers in situ," *Acta Physiol. Scand.*, vol. 70, no. 287, pp. 1–112, 1966.

[3] P. J. Blijham, H. J. T. Laak, H. J. Schelhaas, B. G. M. van Engelen, D. F. Stegeman, and M. J. Zwarts, "Relation between muscle fiber conduction velocity and fiber size in neuromuscular disorders," *J. Appl. Physiol.*, vol. 100, no. 6, pp. 1837–1841, Jun. 2006.

[4] A. Kossev, N. Gantchev, A. Gydikov, Y. Gerasimenko, and P. Christova, "The effect of muscle fiber length change on motor units potentials propagation velocity," *Electromyogr. Clin. Neurophysiol.*, vol. 32, no. 6, pp. 287–294, 1992.

[5] D. Farina, C. Cescon, and R. Merletti, "Influence of anatomical, physical, and detection-system parameters on surface EMG," *Biol. Cybern.*, vol. 86, no. 6, pp. 445–456, Jun. 2002.

[6] W. Li and K. Sakamoto, "The influence of location of electrode on muscle fiber conduction velocity and EMG power spectrum during voluntary isometric contraction measured with surface array electrodes," *Appl. Hum. Sci.*, vol. 15, no. 1, pp. 25–32, 1996.

[7] D. M. Pincivero, R. M. Campy, Y. Salfetnikov, A. Bright, and A. J. Coelho, "Influence of contraction intensity, muscle, and gender on median frequency of the quadriceps femoris," *J. Appl. Physiol.*, vol. 90, no. 3, pp. 804–810, Mar. 2001.

[8] A. Rainoldi, M. Gazzoni, and G. Melchiorri, "Differences in myoelectric manifestations of fatigue in sprinters and long distance runners," *Physiol. Meas.*, vol. 29, no. 3, pp. 331–340, Mar. 2008.

[9] D. Farina, "Counterpoint: Spectral properties of the surface emg do not provide information about motor unit recruitment and muscle fiber type," *J. Appl. Physiol.*, vol. 105, no. 5, pp. 1673–1674, Nov. 2008.

[10] J. M. Wakeling and A. I. Rozitis, "Spectral properties of myoelectric signals from different motor units in the leg extensor muscles," *J. Exp. Biol.*, vol. 207, no. 14, pp. 2519–2528, Jun. 2004.

[11] V. Von Tscharner and B. Goepfert, "Estimation of the interplay between groups of fast and slow muscle fibers of the tibialis anterior and gastrocnemius muscle while running," *J. Electromyogr. Kinesiol.*, vol. 16, no. 2, pp. 188–197, Apr. 2006.

[12] T. W. Boonstra and M. Breakspear, "Neural mechanisms of intermuscular coherence: Implications for the rectification of surface electromyography," *J. Neurophysiol.*, vol. 107, no. 3, pp. 796–807, Feb. 2012.

[13] Á. Costa-García, E. Iáñez, M. Yokoyama, S. Ueda, S. Okajima, and S. Shimoda, "Quantification of high and low sEMG spectral components during sustained isometric contraction," *Physiol. Rep.*, vol. 10, no. 10, May 2022, Art. no. e15296.

[14] A. J. Cruz-Jentoft et al., "Sarcopenia: Revised European consensus on definition and diagnosis," *Age Ageing*, vol. 48, no. 1, pp. 16–31, Jan. 2019.

[15] A. E. Emery, F. Muntoni, and R. Quinlivan, *Duchenne Muscular Dystrophy* (Oxford Monographs on Medical Genetics), no. 67. London, U.K.: Nature Publishing Group, 2015.

[16] A. E. Emery, "The muscular dystrophies," *Lancet*, vol. 359, no. 9307, pp. 687–695, 2002.

[17] P. Turakhia, B. Barrick, and J. Berman, "Patients with neuromuscular disorder," *Med. Clinics North Amer.*, vol. 97, no. 6, pp. 1015–1032, Nov. 2013.

[18] B. Katirji, H. J. Kaminski, and R. L. Ruff, *Neuromuscular Disorders in Clinical Practice*. Berlin, Germany: Springer, 2013.

[19] S. Rahman and M. G. Hanna, "Diagnosis and therapy in neuromuscular disorders: Diagnosis and new treatments in mitochondrial diseases," *J. Neurol., Neurosurg. Psychiatry*, vol. 80, no. 9, pp. 943–953, Sep. 2009.

[20] J. E. Morley, "Sarcopenia: Diagnosis and treatment," *J. Nutrition Health Aging*, vol. 12, no. 7, pp. 452–456, Sep. 2008.

[21] S. Okajima et al., "Quantification of extent of muscle-skin shifting by traversal sEMG analysis using high-density sEMG sensor," in *Proc. IEEE Int. Conf. Cyborg Bionic Syst. (CBS)*, Sep. 2019, pp. 272–277.

[22] D. Stegeman, "Standards for surface electromyography: The European project surface EMG for standards for surface electromyography: The European project surface EMG for non-invasive assessment of muscles (SENIAM)," *Enschede Roessingh Res. Develop.*, vol. 10, pp. 8–12, Jan. 2007.

[23] D. A. Freedman, *Statistical Models: Theory and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[24] Á. Costa-García et al., "Segmentation and averaging of sEMG muscle activations prior to synergy extraction," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3106–3112, Apr. 2020.

[25] N. Gillis, *Nonnegative Matrix Factorization*. Philadelphia, PA, USA: SIAM, 2020.

[26] Y. Zhang and Y. Fang, "A NMF algorithm for blind separation of uncorrelated signals," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit.*, vol. 3, Nov. 2007, pp. 999–1003.

[27] M. Šavc, V. Glaser, J. Kranjec, I. Cikajlo, Z. Matjacic, and A. Holobar, "Comparison of convolutive kernel compensation and non-negative matrix factorization of surface electromyograms," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 10, pp. 1935–1944, Oct. 2018.

[28] E. S. Pearson, "The test of significance for the correlation coefficient," *J. Amer. Stat. Assoc.*, vol. 26, no. 174, pp. 128–134, Jun. 1931.

[29] P. V. Komi and P. Tesch, "EMG frequency spectrum, muscle structure, and fatigue during dynamic contractions in man," *Eur. J. Appl. Physiol. Occupat. Physiol.*, vol. 42, no. 1, pp. 41–50, Sep. 1979.

[30] F. Di Nardo, T. Basili, S. Meletani, and D. Scaradozzi, "Wavelet-based assessment of the muscle-activation frequency range by EMG analysis," *IEEE Access*, vol. 10, pp. 9793–9805, 2022.