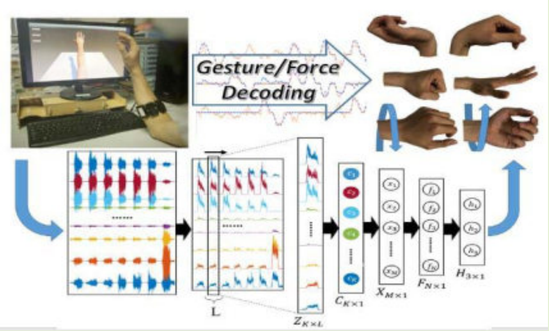
Robust Continuous Hand Motion Recognition Using Wearable Array Myoelectric Sensor

Xuhui Hu, Hong Zeng, *Member, IEEE*, Aiguo Song, *Senior Member, IEEE*, and Dapeng Chen, *Member, IEEE*



***Abstract*—With the advantages of comfortable wearing and outdoor usage, the myoelectric gesture recognition techniques have gained much attention in the field of human-machine interaction (HMI). The purpose of this study is to optimize model structure and transfer generalized features to improve the robustness of myoelectric hand motion decoding. We derived the hand motion recognition framework from the muscle synergy theory, which is formulated as a temporal convolutional (TC) model of array sEMG signals, then a hierarchical myoelectric decoding model was proposed to predict simultaneous and continuous hand motion. The model was trained by the methods of unsupervised low-level feature learning and automated data labeling to minimize training supervision. Extensive experiments on the public sEMG database (17 subjects in Biopatrec) show that the TC model can extract muscle synergy features with higher fidelity ($R^2 = 0.85\pm0.23$) than the traditional instantaneous mixture model, the results of online test demonstrate robust myoelectric decoding on multiple simultaneous and continuous hand motions. More importantly, the analysis of weights visualization shows that the low-level feature representation layer of TC model can be migrated across the individuals, which provides a transferrable feature extraction layer for generalized hand motion decoding.**

***Index Terms*—Muscle synergies, sEMG array, hand motion prediction, generalization, myoelectric decoding model.**

I. Introduction

HUMAN hand, by the unparalleled ability of dexterous manipulation, has attracted large numbers of researchers to explore its principle of neuromuscular interface with the help of sensor technology. Surface electromyography (sEMG) sensor can noninvasively detect the electrophysiological signals of the muscles from the skin. Therefore, the hand motions originated from the co-contraction of the forearm muscles could be recognized by placing an array of sEMG sensors on the skin surface of the forearm. With the advantages of comfortable wearing and outdoor usage, compared to data glove [1], [2] and computer vision [3], the sEMG based gesture recognition techniques have gained much attention in the field of HMI.

The discrete gestures classification based on pattern recognition and continuous hand motions decoding based on regression model are two main research directions of myoelectric control. With the help of deep learning, the discrete gestures classification supports recognizing more gesture categories [4], while the state-of-art continuous hand motion decoding methods realize both the gesture recognition and continuous force or kinematic estimation [5]. From the perspective of realizing natural human-computer interaction, the continuous hand motion decoding method contributes to a more intuitive interaction experience.

Despite decades of efforts have brought about great progress on myoelectric control, there are still some challenges to be solved: 1) Currently it is still difficult to decode dexterous gestures from sEMG signals, and the predicting accuracy decreases with an increase in the number of gestures to be recognized. 2) The relatively consistent and reproducible input features contribute to a reliable prediction for machine learning

Manuscript received May 23, 2021; accepted July 14, 2021. Date of publication July 19, 2021; date of current version September 15, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 91648206, Grant 61673105, and Grant 62003169; and in part by the Jiangsu Key Research and Development Plan under Grant BE2018004-4. The associate editor coordinating the review of this article and approving it for publication was Dr. Varun Bajaj. *(Corresponding author: Aiguo Song.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee of the University, and performed in line with the Declaration of Helsinki.

Xuhui Hu, Hong Zeng, and Aiguo Song are with the State Key Laboratory of Bioelectronics and Jiangsu Key Laboratory of Remote Measurement and Control, School of Instrument Science and Engineering, Southeast University, Nanjing, Jiangsu 210096, China (e-mail: a.g.song@seu.edu.cn).

Dapeng Chen is with the School of Automation, Nanjing University of Information Science and Technology, Nanjing, Jiangsu 210044, China.

This article has supplementary downloadable material available at https://doi.org/10.1109/JSEN.2021.3098120, provided by the authors.

Digital Object Identifier 10.1109/JSEN.2021.3098120

based decoding model, whereas many studies have shown that electrode shifts and long-term usage could shift the input features, degrading the online recognition performance.

To address the above issues, Hahne *et al.* [6] verified the robust recognition of linear regression models for non-training set gestures (i.e., simultaneous gestures). Lin *et al.* [7] and Yang *et al.* [8] proposed the regression model that optimized the model calibration method to improve the predicting accuracy of continuous hand motions. Muceli *et al.* [9] proposed the NMF method that verified the robustness against electrode offsets. He *et al.* [10] proposed the electrode calibration framework PV, which maintains the recognition performance in long time usage scenarios. Currently, most of the myoelectric decoding methods are based on the instantaneous mixture estimation model, where the input is the instantaneous activation signals of sEMG array sensors. This estimation is based on the assumption that the volume conductor from muscle units to each electrode is estimated as a linear transformation, so it may ignore certain latent nonlinear relationships between some complex gestures and corresponding sEMG signals. Besides, Muceli *et al.* [9] pointed out that the prediction of wrist rotation increases the nonlinearity of the model. Therefore, it is important to verify the model robustness when the number of gestures to be predicted increases.

In this paper, we mainly focused on three key issues of myoelectric control for the practical application of HMI: 1) We proposed a novel nonlinear temporal convolutional (TC) model to robustly recognize simultaneous and continuous hand motions that outside the training set. 2) We designed an automated data labeling method to streamline the difficulties in training set collection. 3) To further enhance the generalized recognition across the subjects, we extracted the transferrable low-level representation layer from the TC model. The experimental results show that the TC model can extract muscle synergy features with higher fidelity than the traditional instantaneous mixture model. In addition, the low-level feature representation layer of TC model can be migrated across the individuals, which provides a transferrable feature extraction layer for generalized hand motion decoding.

II. Methods

A. Continuous Myoelectric Decoding Model

The muscle synergy theory, which has been progressively acknowledged over the past decades [11]–[15], shows that the high-dimensional muscle groups are coordinated by low-dimensional functional modules. In this paper, the mapping relationship between the control signals and the sEMG array signals was studied on the basis of the muscle synergy framework, aiming to improve the robustness of myoelectric control by extracting more generalized muscle synergies.

According to the muscle synergy theory, motor unit action potentials (MUAP) transmitted from the spinal cord commonly drive multiple muscles [16], which can be formulated in Equation 1. The independent components among all MUAP can be seen as muscle synergy features $f_n(t)$ with a number of N. Let the activation function of the mth muscle (the sum of all muscle unit action potentials generated from the same motor neural pool) is $x_m(t)$, so it can be seen as a linear representation of N muscle synergy features $[f_1(t), \ldots, f_n(t)]$:

$$x_m(t) = \sum_{n=1}^{N} s_{nm} \cdot f_n(t) \tag{1}$$

where s_{nm} is the weights for different muscle synergies. Since we used non-invasive electrodes attached to the skin of the forearm to acquire sEMG, the signal transmitted from the muscles to the electrode can be seen as a mixture of multiple muscle sources underlying the detecting electrode, and being filtered by tissue and skin conduction [9] (commonly referred to volume conductor), which is expressed in Equation 2:

$$\begin{aligned} z_k(t) &= \sum_{m=1}^{M} \sum_{l=0}^{L} g_{mk}(l) \cdot x_m(t-l) \\ &= \sum_{m=1}^{M} g_{mk}(t) * x_m(t) \end{aligned} \tag{2}$$

It shows that instantaneous sEMG signal acquired by the kth electrode $z_k(t)$ is contributed by all N muscles. Meanwhile, the contribution of each muscle is the convolution between the muscle unit potential sequence with the length of L and volume conductor function $g_{mk}(t)$. Equation 2 depicts an encoding process from $x_m(t)$ to $z_k(t)$, in order to obtain the inverse decoding expression, the Fourier transform is applied to Equation 2, which gives Equation 3:

$$\begin{aligned} z_k(\omega) &= \sum_{m=1}^{M} g_{mk}(\omega) \cdot x_m(\omega) \\ &= [x_1(\omega), \ldots, x_m(\omega)] \cdot \begin{bmatrix} g_{1k}(\omega) \\ g_{2k}(\omega) \\ \vdots \\ g_{mk}(\omega) \end{bmatrix} \end{aligned} \tag{3}$$

where the convolution on the time domain becomes the multiplication on the frequency domain. Due to its linearity property, it can be expressed in a matrix form, as given in Equation 4:

$$[z_1(\omega), \ldots z_k(\omega)] = [x_1(\omega), \ldots x_m(\omega)] \cdot G_{M\times K} \tag{4}$$

$$[x_1(\omega), \ldots x_m(\omega)] = [z_1(\omega), \ldots z_k(\omega)] \cdot G^{+}_{K\times M} \tag{5}$$

It shows that, on the frequency domain, the vector of muscle activation function $[x_1(\omega), \ldots x_m(\omega)]$ can be transformed into sEMG array signals $[z_1(\omega), \ldots z_k(\omega)]$ by matrix $G_{M\times K}$ (Unless otherwise specified, the capital italic symbols indicate matrices and the subscripts indicate their dimensions). Therefore, by computing the pseudo-inverse matrix of G, i.e., G^{+} in Equation 5, $x_m(\omega)$ can be linearly represented by $[z_1(\omega), \ldots z_k(\omega)]$, where the column vector $\left[g'_{1m}(\omega), \ldots g'_{km}(\omega)\right]^{\mathrm{T}}$ in Equation 6 denotes the column vector weights of G^{+} in the mth column. Finally, by applying

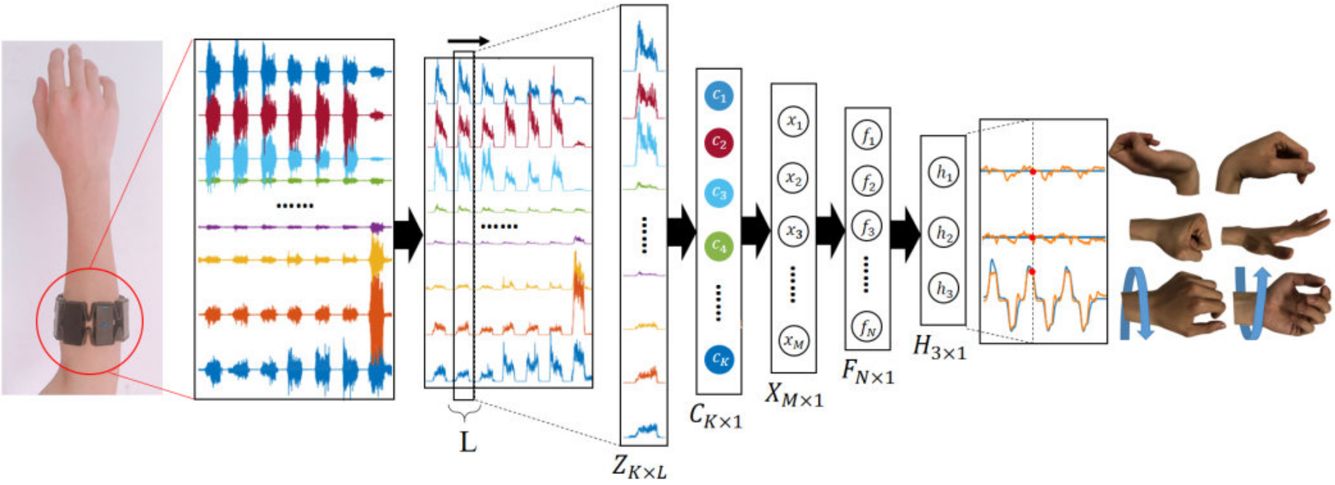


Fig. 1. The framework of temporal convolution based myoelectric decoding model.

Fourier inverse transform on Equation 6, we got Equation 7:

$$x_m(\omega) = [z_1(\omega), \ldots z_k(\omega)] \cdot \begin{bmatrix} g'_{1m}(\omega) \\ g'_{2m}(\omega) \\ \vdots \\ g'_{km}(\omega) \end{bmatrix} = \sum_{k=1}^{K} g'_{km}(\omega) \cdot z_k(\omega) \tag{6}$$

$$x_m(t) = \sum_{k=1}^{K} g'_{km}(t) * z_k(t) \tag{7}$$

Equation 7 shows that the muscle activation function $x_m(t)$ can be obtained by convolving $z_k(t)$ and $g'_{km}(t)$ by channel-wise, and then summing them together. Through combining Equation 7 with Equation 1, the complete decoding model from the sEMG array signals $[z_1(t), \ldots z_k(t)]^T$ to muscle synergy features $[f_1(t), \ldots f_N(t)]^T$ was obtained.

Based on the above formulations, the framework of the proposed myoelectric decoding model can be expressed in Figure 1. In this study, the method of root means square, which is generally estimated to extract the activation signal in literature, was used at first on K channels of sEMG array sensors. During the real-time myoelectric control, a sliding window with a length of L was used to intercept the sequence of sEMG array signals, i.e., $Z_{K\times L}$. The rows of $Z_{K\times L}$ contain the time-domain features and the columns contain the spatial-domain features. According to Equation 6, we performed the convolution operation on the sEMG sequence of each channel (i.e. $C_{K\times 1}$). After that, a linear transformation $X_{M\times 1}$ was performed. Further, $X_{M\times 1}$ was linearly transformed into $F_{N\times 1}$ according to Equation 1. Finally, muscle synergy features were converted into the desired hand motion intentions $H_{3\times 1}$, which represents the number of degrees of freedom (DOFs) to be identified as three.

Through the optimization of the model structure, the instantaneous linear mixing model was transformed into the temporal convolutional model. Due to the combination of temporal features, this method may has the potential to extract more effective muscle synergy features from complex and dexterous gestures. The experimental validation of robust feature extraction and motion recognition is described in Section III. The following subsections introduce the training methods of the temporal convolutional model.

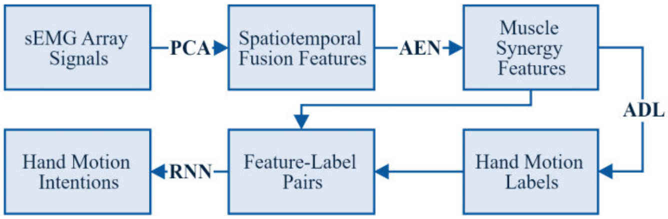


Fig. 2. The training flow chart of myoelectric decoding model.

B. Unsupervised Training for Decoding Model

In order to train the regression model presented in Figure 1, the researchers used various dimensionality reduction (DR) algorithms to extract latent motion intentions such as principal component analysis (PCA) [17], non-negative matrix factorization (NMF) [18], [19], and autoencoder (AEN) [12], [14]. The advantage of these algorithms lies that they can train the feature representation layer, such as $C_{K\times 1}$, $X_{M\times 1}$ or $F_{N\times 1}$, with an unsupervised fashion. However, it is difficult to directly obtain the desired motion intent output, i.e., $H_{3\times 1}$, by self-learning of the machine. As Vujaklija *et al.* [12] pointed out, it required multiple training sessions to obtain the desired decoding model, thus reducing the efficiency of model training. The muscle synergy features can be seen as a weakly labeled sequence [20]–[22], our aim is to determine the correct position of the known hand motion categories in the sequence. To combine the advantages of unsupervised learning with the correct mapping of motion intentions, we designed an automated data labeling method to minimize manual intervention.

The training flow chart of the model is shown in Figure 2. Firstly, a convolution kernel matrix was operated on each myoelectric channel. With the advantages of unsupervised learning and computational efficiency, PCA can be used to obtain c_k by extracting the first principal component of sEMG sequence,

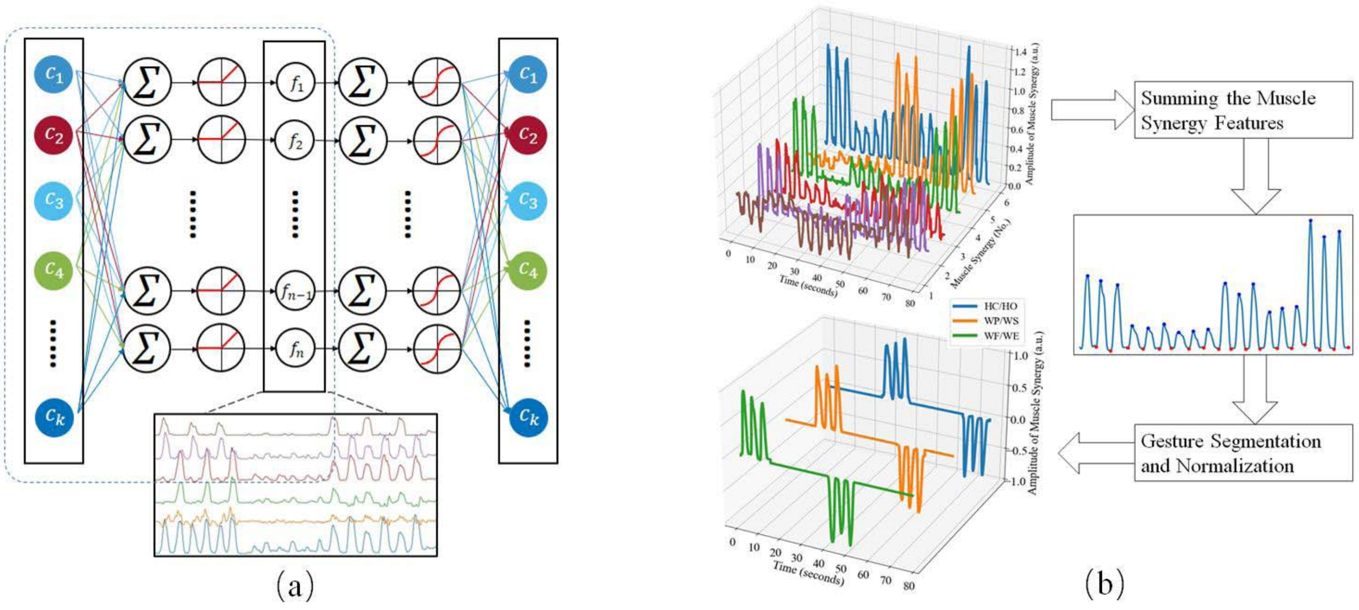

c_1
c_2
c_3
c_4
c_k
f_1
f_2
f_{n-1}
f_n
(a)
Amplitude of Muscle Synergy (a.u.)
Muscle Synergy (No.)
Time (seconds)
Summing the Muscle Synergy Features
Gesture Segmentation and Normalization
HC/HO
WP/WS
WF/WE
(b)

Fig. 3. (a) The structure of auto-encoder. (b) The flow chart of automated data labelling.

which is equivalent to perform a convolution operation on the input signal. Figure 1 shows only one component for each channel, but as the first feature extraction layer, the number of retained components was determined by the explained variance, which should above 90%. Additionally, inspired by the multilayer cylindrical model of volume conductor proposed by Farina *et al.* [23], the volume conductor effect was approximately the same for different electrode channels. Therefore, the PCA matrix was consistent on all channels.

According to Figure 1, the transformation from $C_{K\times 1}$ to $F_{N\times 1}$ was achieved by linear matrix, according to existing studies [24], AEN was adopted to estimate the muscle synergy features. One of the most significant features in AEN is its input neurons are the same as the output neurons, and the number of hidden layer neurons is smaller than the neurons at both ends of AEN. Therefore, the value of the hidden layer can be seen as the potential low-dimensional representation features of the input layer. AEN is also an unsupervised neural network, as is shown in Figure 3(a), the process from the input layer to the hidden layer is called "encoding" (circled with a blue box), and the process from the hidden layer to the output layer is called "decoding". We used the encoding process as a low-level feature representation layer, where the input is the spatiotemporal features $C_{K\times 1}$, and the output is the low-dimensional muscle synergy features $F_{N\times 1}$. To obtain non-negative muscle synergy features, the ReLU function was adopted as the activation function in the encoding process. Since $C_{K\times 1}$ may contain negative components, the Tanh function was adopted in the decoding process to recover negative features of the output neurons. The cross-entropy function was used as the loss function of AEN; the weight matrix of AEN was initialized using the Xavier method; the dropout method was adopted during the iterative training process to prevent over-fitting; the training speed was accelerated by ADAM method and Mini-Batch method. The number of neurons in the hidden layer of AEN was also determined by the explained variance, which should above 85%. As a result, the method of PCA and AEN achieve the low-level feature representation in an unsupervised fashion. Meanwhile, each module can learn the representation independently, making it easier to fine-tune the hyper-parameters on each network layer.

C. Automated Data Labeling

In order to estimate the continuous motions without the external kinematic or kinetic measurements, inspired by the method of force estimation based on motor unit discharge rate [25], we proposed the method of automated data labeling (ADL) [26]. The process of labeling a training set of six gestures by ADL is shown in Figure 3(b), they are wrist flexion (WF), wrist pronation (WP), hand closed (HC), wrist extension (WE), wrist supination (WS) and hand open (HO), each gesture is repeated for three times. ADL takes the sequences of multi-dimensional muscle synergy features (upper left box) as input. We firstly sum these sequences up (right box), the blue sequence represents the oscillation of muscle activation from relaxing to maximum force and back to relaxing. Since we know the order of the collected gestures and the number of repetitions in the data collection procedure, we calculated the start and endpoints of each specific gesture by searching for the minimum (red dots) of this summed sequence and then constructed the gesture label through sequence segmentation.

To decode the simultaneous gestures in real-time, we constructed the gesture label as a three-element list, which denotes the continuous motion of three DOFs controlled by three pairs of antagonistic muscles (i.e. HO and HC, WF and WE, WP and WS). The label value of the rest gesture is zero, and the label value of the two directions within the same DOF is distinguished by the positive or negative value. Since fewer muscle groups are controlling the wrist rotation compared to other DOFs, its peak value in the summed muscle synergy sequence is lower than other gestures. Therefore, we normalized the label value of each DOF into $[-1,1]$ to debias the force level of different DOFs. Finally, the multi-dimensional muscle synergy features extracted by AEN and the motion label generated by ADL constituted the Feature-Label pairs, then they were feed into a regression neural network (RNN) to learn the potential mapping. The RNN was designed with single hidden layer. We adopted ReLU as the activation function, and ADAM method as the solver for weight optimization. The size of the hidden layer and L2 penalty (regularization term) parameter were determined by the preliminary validation test, which was set as 50 neurons and 0.001 respectively.

D. Hypothesis of Robust Myoelectric Decoding

Several studies have reported the robust myoelectric decoding, where the robustness is defined both in terms of reliable recognition on simultaneous gestures outside the training set [6], [9], [19], as well as the general myoelectric decoding that can adapt to different subjects [27]. One of the reasonable explanations for the robust myoelectric decoding is that the low-level representation layer of the decoding model has a robust feature extraction capability, which can stably retain the effective muscle synergy features. Comparing to a deep learning based specific-subject decoding model that solely relies on the terminal-to-terminal estimation, the physiological-inspired hierarchical model can better avoid overfitting to the training set. To verify the above hypothesis, the model robustness analysis was studied to test the generalization extraction ability of the low-level representation layers of TC model, i.e., PCA layer and AEN layer. The specific experimental protocol was described in detail in Section III.

III. Experiments

A. Data Preprocessing

As is shown in Figure 1, the raw sEMG signal was firstly preprocessed to obtain the smoothed sEMG envelope, then it was packaged into a fixed-length sequence. The sampling frequency was decreased to 20 Hz to reduce the redundant training samples. According to the research on the effect of the length of sEMG array sequence to the classification accuracy presented by Betthauser *et al.* [28], we set the sliding time-step as 50ms (equals to sampling frequency) and the length of the window (L in Equation 2) as 20 (one second). The sEMG envelope was normalized from zero to one. Here the normalization factor was consistent on all channels, so that the relative activation variation of different muscle groups during the motion can be retained.

B. Evaluation Criteria

The decoding performance of the model was evaluated at two different levels. The first is the commonly used predicting accuracy, which is evaluated by the coefficient of determination (denoted as $R^{\mathbf{2}}$) on each gesture. The two measured variables in $R^{\mathbf{2}}$ were the expected labels estimated by ADL and the predicted value of the myoelectric decoding model. The second is the reconstruction goodness of muscle synergy. Since the decoding model was trained by the unsupervised learning algorithm, the corresponding reconstruction matrix was trained along with the decoding matrix. Here the reconstructed sEMG signals from the muscle synergy features and the sEMG signals before the reconstruction were compared using $\overline{R^2}$, which is the averaged coefficient of determination among subjects, gestures and channels. Once the extracted muscle synergy features could restore the original input with higher fidelity, it indicates that most of the latent features were preserved for the higher level of motion intention mapping. Whereas in the case of poor reconstruction, some of the latent features may be lost during the transmission of the low-level layer, which degrades the robustness of feature extraction.

C. Experimental Protocol

The experiments consist of both offline and online tests. In the offline test, we used a publicly available dataset "6mov8chUFS" from "Biopatrec" [29] to fully validate the robustness of the proposed decoding model, which contains six basic hand and wrist motions from 17 intact subjects, as well as 20 possible simultaneous gestures of the basic motions. In the online test, we used MYO (Thalmic Labs), a commercial myoelectric armband with eight bipolar dry electrode channels, to perform real-time continuous hand motion recognition. The detailed description of the experimental protocol is shown in Table I. The "Session" column indicates the purpose of each test, the "Comparison" lists all the items to be compared, "Details" complements the requirements of model parameters and training methods, "Data" specifies the source of the training set and test set, and "Indicator" specifies the evaluation criteria.

The difference in the myoelectric decoding model was tested in Session I, it was assumed that the proposed temporal convolutional model (TC) could extract muscle synergy features with higher fidelity than the instantaneous linear mixing model (IM). The weights of TC and IM were trained in a specific-subject manner, and the training set only contains single DOF gestures. In "6mov8chUFS", all the subjects performed three repetitions on each gesture. Unless otherwise specified, we took the sEMG array sequences of the first two rounds of single DOF gestures as the training set, and took the last round of single DOF gesture and all the simultaneous gestures as the test set. Since the differences between the two decoding models mainly focused on the feature extraction layer (differ from feature mapping of RNN), the reconstruction goodness of muscle synergy was adopted as the evaluation criteria, which was tested on both single DOF and simultaneous gestures

The generalization of the unsupervised feature extraction layers was tested in Session II, where the data from one of the 17 subjects was selected as the training set and the data from the other 16 subjects were used as the test set. Since it was aimed at evaluating the process of muscle synergy extraction, so the same evaluation approach as in Session I was adopted.

In Session III, 10 able-bodied subjects participated in the online test, the informed consent form was obtained from all the participated subjects, all the experiments were approved by the Ethical Committee of the university and conformed to the Declaration of Helsinki. The myoelectric decoding

TABLE I
DETAILS OF EXPERIMENTAL PROTOCOL

	Session		Comparison	Details	Data	Indicator
Offline Tests	I.	Model	(a)Instantaneous Linear (b)Temporal Convolution	Specific-subject training	Biopatrec	Reconstruction of Muscle Synergy
	II.	Training Set	(a) Specific-subject (b) General-subject	The same *framework* as in the previous test		
Online Tests	III.	Motion Decoding	(a) Instantaneous Linear (b) Temporal Convolution	The same *training set* as in the previous test	MYO Armband	Predicting Accuracy

TABLE II
DESCRIPTIVE STATISTICS OF RECONSTRUCTION GOODNESS

	Number of Components (Compression Ratio)	Minimum	Maximum	Range	Mean($\overline{R^2}$)	Std. Deviation
IM: NMF	4 (50.0%)	-39.76	1.00	40.76	0.45	1.764
	5 (37.5%)	-45.62	1.00	46.62	0.68	1.474
	6 (25.0%)	-23.11	1.00	24.11	0.86	0.727
	7 (13.5%)	-15.72	1.00	16.72	0.93	0.435
1^{st} Layer of TC: PCA	1 (95.0%)	-2.39	0.88	3.27	0.68	0.165
	2 (90.0%)	-1.82	0.97	2.79	0.85	0.143
	3 (85.0%)	0.25	0.98	0.74	0.90	0.091
	4 (80.0%)	0.35	0.99	0.64	0.93	0.071
2^{nd} Layer of TC: AEN (1^{st} CR=80%)	16 (50.0%)	-19.00	0.98	19.98	0.70	0.706
	20 (37.5%)	-37.60	0.98	38.58	0.78	0.768
	24 (25%)	-5.72	0.99	6.71	0.83	0.293
	28 (13.5%)	-4.92	0.99	5.92	0.86	0.233

model was designed to recognize six gestures originated from three DOFs of the wrist and hand, they were wrist flexion (WF), wrist extension (WE), wrist pronation (WP) and wrist supination (WS). The low-level representation layer of the temporal convolutional decoding model was migrated from the result of Session II. The training set of RNN contained three repetitions on each single DOF gestures, and the online predicting accuracy was tested on both single DOF and simultaneous gestures.

IV. RESULTS AND DISCUSSIONS

A. Results of Model Comparison

Firstly, we compared the performance of muscle synergy extraction using Instantaneous Mixing Model (IM) and Temporal Convolutional Model (TC) respectively. According to the experimental protocol described in Section III.C, the sEMG sequences were divided by different subjects, gestures, and channels, then R^2 was calculated at each interval. Therefore, the sample capacity for each treatment group was 3536 ($17 \times 26 \times 8$), the results of statistics, such as mean ($\overline{R^2}$) and standard deviation (SD) are shown in Table II.

For the sEMG array sensor with eight channels, the compression ratio of IM is limited (ranging from one to seven). We evaluated the compression range from seven components compression ratio), as is shown in the first row of Table II. The statistic results show that even with the lowest achievable compression ratio (13.5%), IM achieves minimum reconstruct-ion goodness of -15.72 and SD of 0.4352, indicating that it is difficult to obtain the stable reconstruction performance across all gestures and channels as more gestures needed to be recognized.

The second row of Table II shows the reconstruction goodness of the sEMG sequence intervals about the number of the retained time-domain components extracted by PCA. The results show that when the compression ratio is at 80% (i.e. a sliding window sequence of length 20 is compressed to four components), $\overline{R^2}$ is close to IM at 0.9302, but the minimum and SD are better than IM. Further, we used the same data compression ratio as in IM to compare the reconstruction goodness of the muscle synergy features extracted by AEN in TC, the statistics results is shown in the third row of Table II, and Figure 4(a) was plotted based on the above statistical results. It is shown that the reconstructed goodness of IM decreases significantly with increasing compression ratio, and the mean value of IM are close to TC when the compression ratio is less than 75%, but the variance of IM is significantly larger than that of TC. We set the compression ratio of IM to 13.5% and set the compression ratio of PCA and AEN in TC to 80% and 13.5% respectively. Figure 4(b) shows the reconstruction goodness of two models across each channel, where the first channel of the sEMG array sensor was consistently placed along the extensor carpi ulnaris and

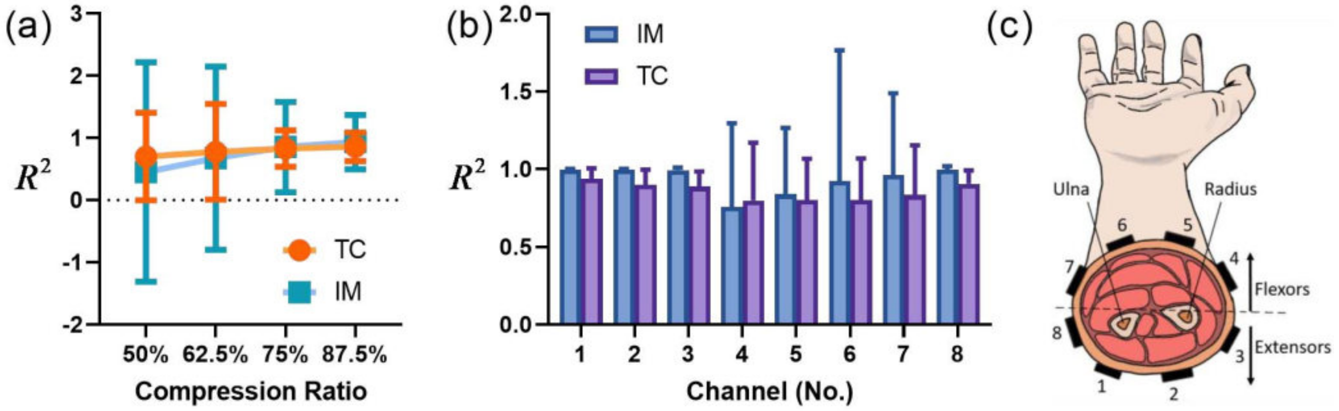


Fig. 4. Model comparison based on the Biopatrec dataset. (a) Comparison of reconstruction goodness using the instantaneous mixing model (IM) and the temporal convolutional model (TC) at different compression ratios. (b) The comparison of the reconstruction goodness on different channels at compression ratio of 87.5%. (c) The cross section of the forearm. The black rectangles denote the distribution of the sEMG array sensors in "6mov8chUFS", the index of the electrode correspond to the horizontal coordinates in the left panel.

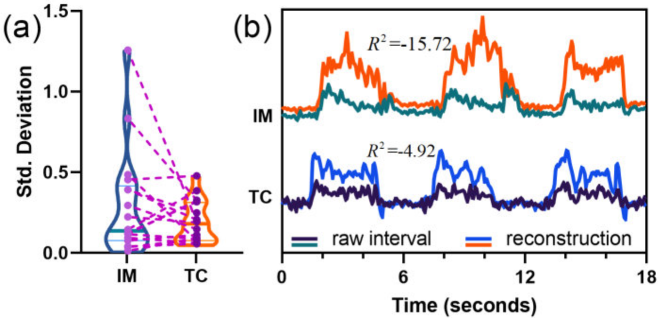


Fig. 5. (a) The distribution of the standard deviation of the reconstruction goodness at the compression ratio of 87.5%. (b) The intervals with the worst reconstruction goodness by IM and TC. The upper interval is the reconstruction on the 4th of channel, 2nd of gesture and 3rd of subject by TC; the lower interval is on the 6th of channel, 26nd of gesture and 7rd of subject by IM.

the rest equally spaced following the lateral direction around the forearm [24]. The results show that, in IM, the SD of reconstruction goodness on the side of extensor muscles is significantly larger than that of flexor muscles. Whereas in TC, the differences of reconstruction goodness among channels are not significant, indicating that all the sEMG sensor makes a balanced contribution to the extraction of muscle synergy features. The above results indicate that TC model is more robust to extract effective muscle synergy features from different gestures and channels than IM, demonstrating statistically superiority among all subjects.

Figure 5(a) depicts the distribution of the SD of the reconstruction goodness over all subjects, it is shown that the fluctuations of reconstruction goodness in IM model commonly appear among subjects, where six out of seventeen subjects had a decline of SD that more than 0.1 when adopting from IM to TC. Among them, three subjects had a decline of more than 0.4, whose SD in IM was higher than the maximum of SD in TC. The intervals with the worst reconstruction goodness of IM ($R^2 = -15.72$) and TC ($R^2 = -4.92$) were visualized in Figure 5(b), which shows that IM causes a greater baseline shift than TC. It is also shown that the amplitude of these sEMG intervals is less than that in other channels and gestures (indicating fewer MUAPs are recruited). However, the muscle synergy features extracted by two methods enhance the amplitude of the activated intervals, where TC can better recover the trend of the activation signal than IM.

The results of model comparison show that after joining more single DOF gestures and simultaneous gestures, IM has a significant decline of the reconstruction goodness on certain gestures and channels. However, TC can explain the variance of sEMG array signal in a more robust and high-fidelity way, contributing to a higher level of regularization for simultaneous gestures recognition.

B. Validation on Robustness Hypothesis

The weights similarity of the 17 specific-training models was analyzed to validate the robustness hypothesis mentioned in Section II.D. The weight distribution of the PCA layer is shown in Figure 6(a). The DR matrix of PCA (dimension: 20 × 4) was divided into four-column vectors, representing four convolution kernel function for extracting the time-domain features ($g'_{km}(t)$ in Equation 7).

It is found that the weight vectors of the third and fourth temporal component extracted by certain subjects may opposite to others (e.g., the absolute values of weights are similar, but the signs are opposite). This difference can be seen as extracting features that negative to others, but it can still reconstruct the sEMG by changing the sign again. As a result, if the weight matrices trained by two subjects are only opposite in sign, their weights can still migrate to each other. The sign of weight matrices was aligned and plotted in Figure 6(a), where the violet strip is the error band, and the solid purple line is the mean of the weights. The results indicate that the weight matrix of the PCA layer has a high similarity across the subject.

Then the similarity of the encoding layer of AEN was analyzed, we reshaped the encoding matrix into one column vector to compute cosine similarity, the cosine of the two-column vectors equals one when they are the same. Figure 6(b) shows the similarity matrix of the encoding layer, the element of

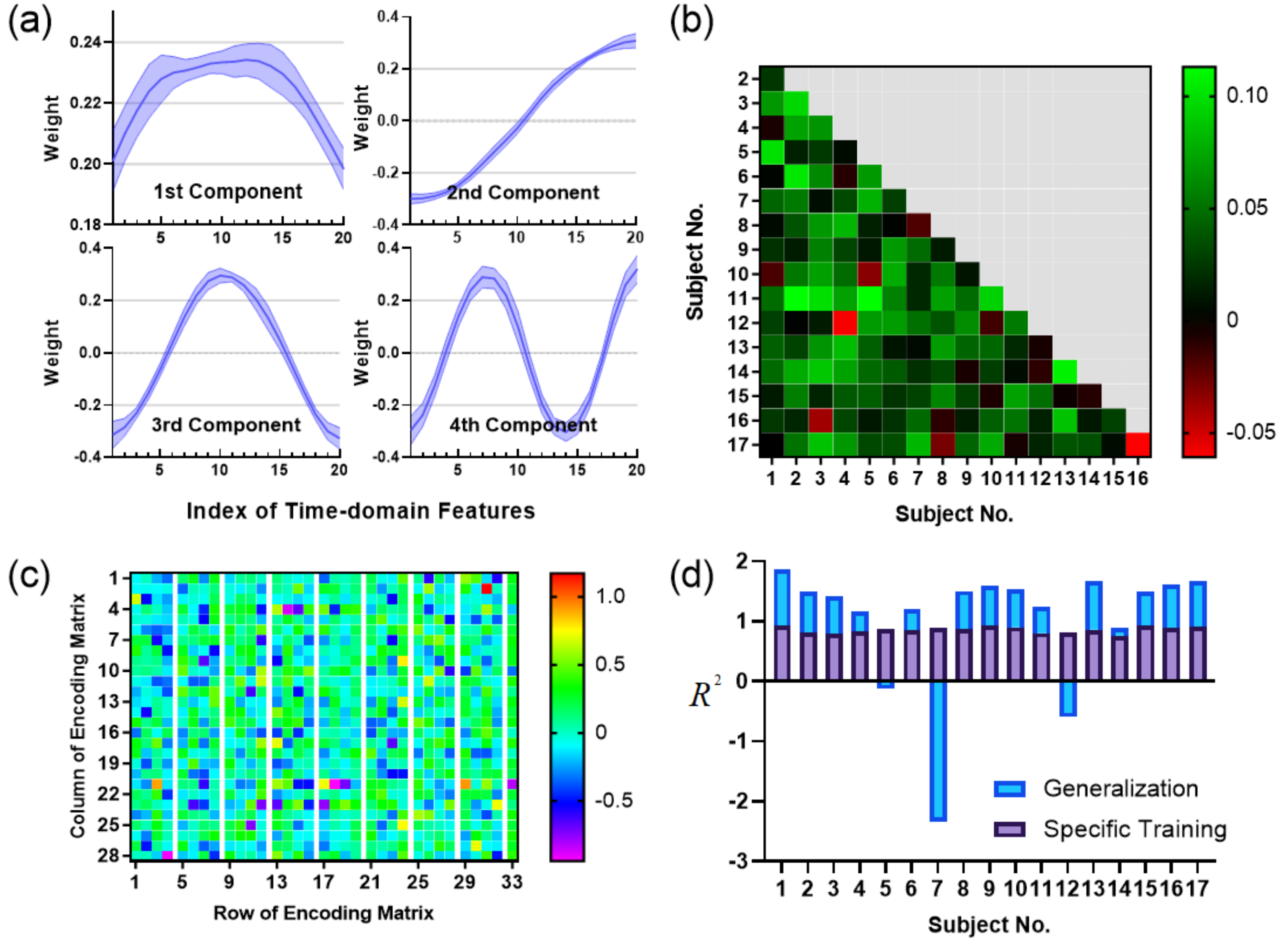


Fig. 6. (a) The weights distribution of the PCA layer among 17 subjects in TC. (b) The similarity matrix of the encoding layer of AEN in TC. (c) The encoding matrix of subject I. (d) The comparison of reconstruction goodness between the generalization model and the specific training model.

the matrix denotes the cosine similarity between two subjects, whereas it is found that the similarity of the encoding matrix across 17 subjects is less than 0.1. The encoding matrix of subject I is shown in Figure 6(c), the linear transformation matrix with the dimension of 28 × 32 is horizontally stacked on the left side, and the bias weight vector is on the right side with the dimension of 28 × 1. The white gap distinguishes eight weighting matrices, which were used to extract muscle synergies from different channels. It is difficult to find intuitive similarity between the weighting matrices of the different channels from Figure 6(c), indicating that the process of muscle synergy extraction varies across different channels

To validate the generalization ability of the low-level feature representation layers, as is described in Session II of the experimental protocol, the reconstruction goodness of muscle synergy is shown in Figure 6(d). The results show that after migrating the weight matrix, most of the subjects experience the reconstruction distortion. As a result, the PCA layer of TC model can extract generalized features across subjects, whereas the encoding layer of AEN still needs to perform specific training to ensure high-fidelity extraction of muscle synergy features.

C. Online Myoelectric Decoding Results

Based on the results of the generalization experiments in subsection B, we migrated the weight matrix of the PCA layer across all the subjects, while AEN and RNN continued with individual training, and the training set only contained single DOF gestures.

In the online test phase, ten subjects sequentially completed six single DOF gestures and eight commonly used synchronous gestures, the predicted sequence of one of the subjects whose recognition rate was at the average level is shown in Figure 7(a). As a comparison, we trained a IM model with the same training set, the statistic results of online test are shown in Figure 8, all the gestures were categorized by DOFs, they were (1) hand open and close (HO/HC) (2) wrist flexion and extension (WF/WE) (3) wrist pronation and supination (WP/WS). The statistical results show that in TC, the predicting accuracy of HO/HC is 0.65±0.23, WF/WE is 0.81±0.17, WP/WS is 0.31±0.58. In IM, HO/HC is 0.47±0.24, WF/WE is 0.56±0.40, WP/WS is 0.03±0.66.

D. Discussions

In this paper, a temporal convolutional model was proposed for continuous hand motion recognition by array sEMG sensor.

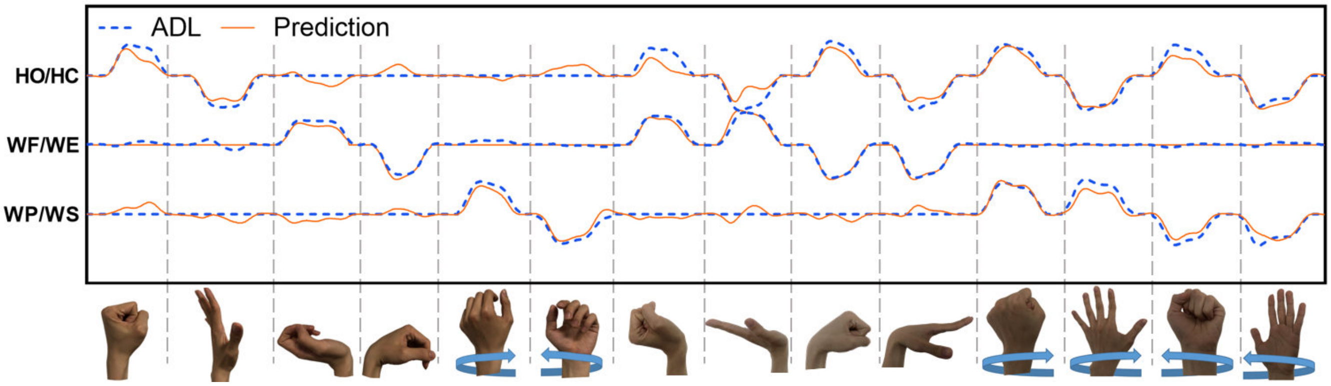


Fig. 7. Real-time continuous hand motion regression sequence of three DOFs.

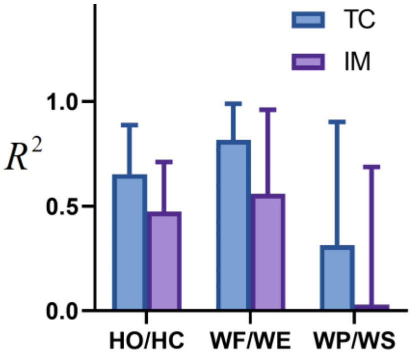


Fig. 8. Fitting accuracy of continuous hand motions measured on R^2.

The experimental results demonstrate that TC has a more balanced restoration on the array sEMG signals on both the flexor and extensor side of the forearm. When TC and IM adopted the same compression ratio, the standard deviation of the reconstruction goodness in TC is significantly smaller than that in IM, indicating that TC model improves the robust feature extraction and recognition for simultaneous gestures that outside the training set. Meanwhile, we extracted a transferable feature representation layer from the TC model, which helps to build a generalized myoelectric decoding model.

From an intuitive point of view, the input features of IM only contain spatial information, whereas the first layer of TC uses PCA to extend the input to spatiotemporal relevant features. Comparing to directly feed the spatiotemporal sEMG image into an end-to-end neural network, TC significantly reduces the model complexity and the training cost of the decoding model. Meanwhile, Table II shows that after the dimensionality reduction processing by PCA and AEN, the reconstruction goodness was not significantly reduced, indicating that the latent features were effectively retained.

In the experiment of model comparison, it is found that a significant distortion occurred in IM model when reconstructing sEMG array signals from the extracted muscle synergy features. As is shown in Figure 4(b), the reconstruction on the side of the extensor muscles appears over-fitted, but on the flexor side, it appears significantly under-fitted, which eventually produces a large reconstruction distortion. By fusing the time-domain features, the reconstruction goodness of TC model is limited within a smaller range, improving the regularization level of the model.

In the online experiment, the decoding accuracy of the temporal convolutional model shows superiority compared to the instantaneous mixture model on both single DOF and simultaneous gestures. However, figure 7(a) also shows crosstalk on decoding single DOF gestures, which may result from the underfitting of the decoding model or the involuntary activation of multiple gestures simultaneously during the training phase. Our future work will continue to focus on optimizing the training efficiency of the model and improving the generalized feature extraction of the decoding model.

V. Conclusion

We presented a novel hierarchical myoelectric decoding model for continuous hand motion recognition by sEMG array sensor. The latent hand motion signal was formulated into the temporal convolutional model with respect to array sEMG signals according to the muscle synergy theory. With the method of unsupervised low-level feature learning and automated data labelling, the model can match the continuous motion label to the muscle synergy features with minimum supervision. The experimental results show that the proposed framework is able to improve the muscle synergy extraction with high level of regularization. More importantly, the analysis on weights visualization shows that the low-level feature representation layer can be shared across individuals, which enhances the capability of continuous and simultaneous hand motion decoding.

References

[1] Y. Zhang *et al.*, "Static and dynamic human arm/hand gesture capturing and recognition via multiinformation fusion of flexible strain sensors," *IEEE Sensors J.*, vol. 20, no. 12, pp. 6450–6459, Jun. 2020.

[2] X. Zhang, Z. Yang, T. Chen, D. Chen, and M.-C. Huang, "Cooperative sensing and wearable computing for sequential hand gesture recognition," *IEEE Sensors J.*, vol. 19, no. 14, pp. 5775–5783, Jul. 2019.

[3] L. Baraldi, F. Paci, G. Serra, L. Benini, and R. Cucchiara, "Gesture recognition using wearable vision sensors to enhance visitors' museum experiences," *IEEE Sensors J.*, vol. 15, no. 5, pp. 2705–2714, May 2015.

[4] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, Nov. 2016, Art. no. 36571.

[5] D. Farina *et al.*, "Man/machine interface based on the discharge timings of spinal motor neurons after targeted muscle reinnervation," *Nature Biomed. Eng.*, vol. 1, no. 2, p. 25, Feb. 2017.

[6] J. M. Hahne *et al.*, "Linear and nonlinear regression techniques for simultaneous and proportional myoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 2, pp. 269–279, Mar. 2014.

[7] C. Lin, B. Wang, N. Jiang, and D. Farina, "Robust extraction of basis functions for simultaneous and proportional myoelectric control via sparse non-negative matrix factorization," *J. Neural Eng.*, vol. 15, no. 2, Apr. 2018, Art. no. 026017.

[8] D. Yang, J. Li, X. Zhang, and H. Liu, "Simultaneous estimation of 2-DOF wrist movements based on constrained non-negative matrix factorization and Hadamard product," *Biomed. Signal Process. Control*, vol. 56, Feb. 2020, Art. no. 101729.

[9] S. Muceli, N. Jiang, and D. Farina, "Extracting signals robust to electrode number and shift for online simultaneous and proportional myoelectric control by factorization algorithms," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 623–633, May 2014.

[10] J. He, X. Sheng, X. Zhu, and N. Jiang, "A novel framework based on position verification for robust myoelectric control against sensor shift," *IEEE Sensors J.*, vol. 19, no. 21, pp. 9859–9868, Nov. 2019.

[11] N. Jiang, K. B. Englehart, and P. A. Parker, "Extracting simultaneous and proportional neural control information for multiple-DOF prostheses from the surface electromyographic signal," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1070–1080, Apr. 2009.

[12] I. Vujaklija, V. Shalchyan, E. N. Kamavuako, N. Jiang, H. R. Marateb, and D. Farina, "Online mapping of EMG signals into kinematics by autoencoding," *J. NeuroEng. Rehabil.*, vol. 15, no. 1, p. 21, Mar. 2018.

[13] N. M. Cole and A. B. Ajiboye, "Muscle synergies for predicting non-isometric complex hand function for commanding FES neuroprosthetic hand systems," *J. Neural Eng.*, vol. 16, no. 5, Aug. 2019, Art. no. 056018.

[14] Y. Yu, C. Chen, X. Sheng, and X. Zhu, "Multi-DoF continuous estimation for wrist torques using stacked autoencoder," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101733.

[15] A. Furui *et al.*, "A myoelectric prosthetic hand with muscle synergy–based motion determination and impedance model-based biomimetic control," *Sci. Robot.*, vol. 4, no. 31, Jun. 2019, Art. no. eaaw6339.

[16] A. d'Avella, A. Portone, L. Fernandez, and F. Lacquaniti, "Control of fast-reaching movements by muscle synergy combinations," *J. Neurosci.*, vol. 26, no. 30, pp. 7791–7810, Jul. 2006.

[17] L. J. Hargrove, G. Li, K. B. Englehart, and B. S. Hudgins, "Principal components analysis preprocessing for improved classification accuracies in pattern-recognition-based myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1407–1414, May 2009.

[18] F. E. Zajac, "Muscle and tendon: Properties, models, scaling, and application to biomechanics and motor control," *Crit. Rev. Biomed. Eng.*, vol. 17, no. 4, pp. 359–411, 1989.

[19] N. Jiang, H. Rehbaum, I. Vujaklija, B. Graimann, and D. Farina, "Intuitive, online, simultaneous, and proportional myoelectric control over two degrees-of-freedom in upper limb amputees," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 501–510, May 2014.

[20] K. Wang, J. He, and L. Zhang, "Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors," *IEEE Sensors J.*, vol. 19, no. 7, pp. 7598–7604, Sep. 2019.

[21] Q. Teng, K. Wang, L. Zhang, and J. He, "The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition," *IEEE Sensors J.*, vol. 20, no. 13, pp. 7265–7274, Jul. 2020.

[22] Y. Tang, Q. Teng, L. Zhang, F. Min, and J. He, "Layer-wise training convolutional neural networks with smaller filters for human activity recognition using wearable sensors," *IEEE Sensors J.*, vol. 21, no. 1, pp. 581–592, Jan. 2021.

[23] D. Farina, L. Mesin, S. Martina, and R. Merletti, "A surface EMG generation model with multilayer cylindrical description of the volume conductor," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 3, pp. 415–426, Mar. 2004.

[24] M. Z. ur Rehman *et al.*, "Stacked sparse autoencoders for EMG-based classification of hand motions: A comparative multi day analyses between surface and intramuscular EMG," *Appl. Sci.*, vol. 8, no. 7, p. 1126, Jul. 2018.

[25] C. Dai, Y. Zheng, and X. Hu, "Estimation of muscle force based on neural drive in a hemispheric stroke survivor," *Frontiers Neurol.*, vol. 9, p. 187, Mar. 2018.

[26] X. Hu, H. Zeng, D. Chen, J. Zhu, and A. Song, "Real-time continuous hand motion myoelectric decoding by automated data labeling," in *Proc. Int. Conf. Robot. Automat. (ICRA)*, Paris, France, May 2020, pp. 6951–6957.

[27] W. Yang, D. Yang, Y. Liu, and H. Liu, "Decoding simultaneous multi-DOF wrist movements from raw EMG signals using a convolutional neural network," *IEEE Trans. Human-Mach. Syst.*, vol. 49, no. 5, pp. 411–420, Oct. 2019.

[28] J. L. Betthauser *et al.*, "Stable responsive EMG sequence prediction and adaptive reinforcement with temporal convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 6, pp. 1707–1717, Jun. 2020.

[29] M. Ortiz-Catalan, R. Brånemark, and B. Håkansson, "BioPatRec: A modular research platform for the control of artificial limbs based on pattern recognition algorithms," *Sour. Code Biol. Med.*, vol. 8, p. 11, Apr. 2013.

Xuhui Hu received the B.S. degree in electrical engineering from the Changshu Institute of Technology, Suzhou, China, in 2016. He is currently pursuing the Ph.D. degree with the School of Instrument Science and Engineering, Southeast University, Nanjing, China. His research interests include human–computer interaction and bionic prostheses.

Hong Zeng (Member, IEEE) received the Ph.D. degree in computer science from Hong Kong Baptist University, Hong Kong, in 2010. He is currently an Associate Professor with the Robot Sensing and Control Technology Laboratory, School of Instrument Science and Engineering, Southeast University, Nanjing, China. His current research interests include biorobot/biomechatronic interfaces and cortically coupled human–machine collaboration.

Aiguo Song (Senior Member, IEEE) received the B.S. degree in automatic control and the M.S. degree in measurement and control from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1990 and 1993, respectively, and the Ph.D. degree in measurement and control from Southeast University, Nanjing, in 1996. He is currently a Professor with the Intelligent Information Processing Laboratory, Southeast University. His research interests include haptic display, robot tactile sensor, and telerehabilitation robot.

Dapeng Chen (Member, IEEE) received the B.S. degree in electrical engineering and automation from the Anhui University of Science and Technology, Huainan, China, in 2011, and the Ph.D. degree in instrumentation science and technology from Southeast University, Nanjing, China, in 2019. He was a Visiting Scholar with the Intelligent, Multimedia, and Interactive Systems Laboratory, University of North Carolina at Charlotte, Charlotte, from 2016 to 2017. He is currently a Lecturer with the School of Automation, Nanjing University of Information Science and Technology, Nanjing. His research interests include haptic display, haptic device, and human–computer interaction.