Reading Your Actions: Learning Generalizable Action Representations via Pre-training AEMG

Supplementary Material

A. More Details about AEMG

A.1. Detailed Architecture of Neuro-Syntax Transformer(NST)

The NST serves as the backbone for both the VQ codebook learning and the Masked EMG Modeling (MEM). It follows a standard Transformer Encoder architecture but is augmented with our proposed Spatio-Temporal embeddings. The detailed hyperparameter configuration used in our AEMG-Base and AEMG-Large models is listed in Table 1.

Table 1. Detailed configurations of NST variants.

Hyperparameter	**AEMG-Base**	**AEMG-Large**
Hidden Dimension (D)	256	512
Feedforward Dim (D_{FFN})	1024	2048
Num. Layers (N_{layers})	6	12
Num. Attention Heads	8	16
Dropout Rate	0.1	0.1
Max Sequence Length (l_{max})	256	256
Vocabulary Size (k)	8192	8192
Parameter Count	$\approx$ 5.2M	$\approx$ 46.5M

A.2. Formal Formulation of Gradient Estimation

A critical challenge in optimizing the NCT is the non-differentiable nature of the Vector Quantization (VQ) step. To enable end-to-end training, we employ the Straight-Through Estimator (STE).

Formally, let $z_e(x)$ be the continuous output of the encoder and $z_q(x)$ be the quantized discrete latent vector. During the **forward pass**, we perform a standard nearest neighbor search over the codebook $\mathcal{V} = \{\mathbf{v}_j\}_{j=1}^{k}$:

$$z_q(x) = \mathbf{v}_k, \quad \text{where } k = \arg\min_j \|z_e(x) - \mathbf{v}_j\|_2 \tag{1}$$

During the **backward pass**, the gradient of the loss $\mathcal{L}$ with respect to the quantization input $z_e(x)$ is approximated by copying the gradient from the decoder input $z_q(x)$:

$$\frac{\partial \mathcal{L}}{\partial z_e(x)} \approx \frac{\partial \mathcal{L}}{\partial z_q(x)} \tag{2}$$

This approximation allows gradients to flow through the bottleneck, updating the encoder parameters despite the non-differentiable $\arg\min$ operation.

To prevent codebook collapse and enforce alignment between the encoder output and the codebook vectors, we minimize the following composite objective:

$$\mathcal{L}_{total} = \mathcal{L}_{rec}(x, \hat{x}) + \|sg[z_e(x)] - \mathbf{v}\|_2^2 + \beta\|z_e(x) - sg[\mathbf{v}]\|_2^2 \tag{3}$$

where $sg[\cdot]$ denotes the stop-gradient operator, and $\beta = 0.25$ is the commitment cost coefficient.

B. Details of Datasets and Preprocessing

In this section, we provide a detailed overview of the heterogeneous EMG datasets utilized in our work. To train the generalization capability of the proposed framework, we aggregated data from multiple sources, covering a wide range of acquisition devices, electrode configurations, and sampling protocols.

B.1. EMG Acquisition Devices

The datasets included in this study are collected using a variety of commercial and research-grade EMG devices, ensuring significant diversity in signal characteristics:

Myo Armband. Used in the Myo Dataset, UCLEMG, and Ninapro DB5. This consumer-grade device consists of 8 dry stainless steel electrodes arranged in a bracelet. Due to the dry electrode interface and wireless low-power design, it typically exhibits higher skin-electrode impedance and operates at a relatively low sampling rate of 200 Hz [16].

Delsys Trigno. Employed in Ninapro DB2 and DB3. These are individually placed wireless bipolar sensors using dry silver electrodes, providing high-fidelity signals with a sampling rate of 2000 Hz [1].

High-Density Arrays. Used in datasets like Capgmyo . These devices utilize grid arrays (e.g., 8×16 or similar configurations) with conductive gel or hydrogel interfaces, capturing high-resolution spatiotemporal muscle activation patterns [5].

B.2. Dataset Specifications

Figure 1 illustrates the electrode configurations across datasets and the generalization tasks for EMG pre-training. The left panel categorizes electrode types into dry and wet electrodes[15], with further classification into individually placed electrodes and electrode arrays. The middle panel details the specific electrode configurations and placements for each of the eight EMG datasets, including Ninapro DB3 which contains subjects with transradial amputations (TA)[8, 10, 15]. The right panel demonstrates how EMG sensors detect electromyographic signals from the arm, revealing distinct gesture-specific patterns of muscle activation.

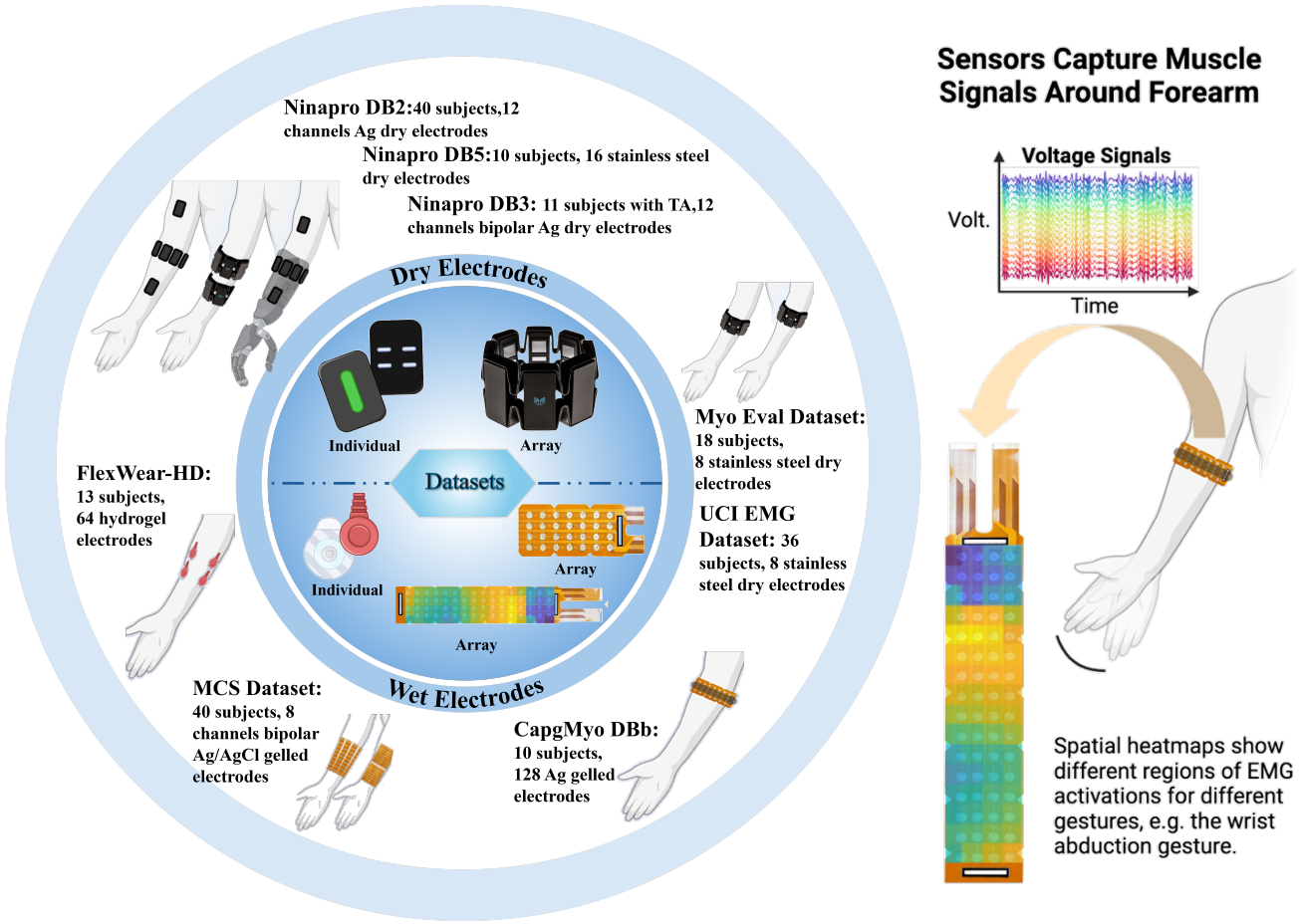

Figure 1. EMG Pre-training Datasets

Table 2. Summary of the heterogeneous EMG datasets used in the pre-training. The table details the sensor type, channel configuration, subject count, and signal characteristics for each dataset.

Dataset	Device Type	Channels	Subjects	Gestures(Classes)	Sampling Freq(Hz)
Capgmyo[5]	Gelled	128	10	8	1000
Ninapro DB2[2]	Dry Bipolar	12	40	10	2000
Ninapro DB3[2]	Dry Bipolar	12	11	10	2000
Myo Dataset[4]	Dry	8	18	7	200
UCI EMG[9]	Dry	8	36	6	1000
Ninapro DB5[12]	Dry	16	10	10	200
MCS[11]	Gelled Bipolar	4	40	7	2000
FlexWear-HD[15]	Hydrogel	64	13	10	4000
Total	**Mixed**	**4–128**	**178**	**Mixed**	**200–4000**

Note: The number of gestures listed refers to the subset used for consistent cross-dataset evaluation, focusing on common wrist and finger movements as described in prior works.

Table 2 summarizes the key specifications of the 8 datasets included in our pre-training corpus. The collection comprises data from over 170 subjects with channel counts ranging from 4 to 128 and sampling frequencies spanning 200 Hz to 4000 Hz[2, 4, 5, 9, 11, 12, 15].

As detailed in the supplementary material of our work, we standardized the data segmentation process. Following standard practices in the literature, samples were generated

Figure 2. EMG Pre-training gestures

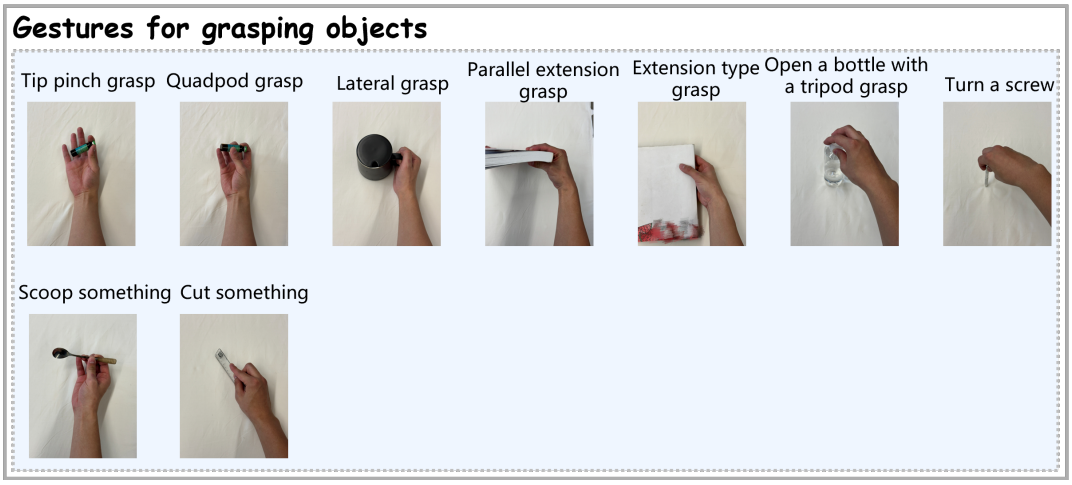

Figure 3. EMG Pre-training gestures(2)

using overlapping sliding windows. Specifically, a window length of 250 ms was selected to balance classification performance with real-time responsiveness (delays under 300 ms are generally acceptable for prosthetic control). A step length of 50 ms was used for most datasets to maximize data density, while a larger step length was applied to high-density datasets to maintain computational tractability.

To further demonstrate the diversity and complexity of the pre-training data, we visualize a subset of the gesture vocabulary in Figure 2 and Figure 3 . The dataset encompasses a wide spectrum of hand movements, ranging from basic single-finger articulations (e.g., index flexion, thumb adduction) to complex multi-finger synergies (e.g., extension of index, ring, and middle fingers). Furthermore, it includes intricate grasping patterns required for object manipulation, such as the precision grip for small objects (e.g., tip pinch, writing tripod grasp) and power grasps for larger items (e.g., medium wrap, power sphere grasp). This rich variety of gesture types ensures that the AEMG framework learns robust and generalizable representations capable of decoding fine-grained motor intentions across diverse functional tasks.

B.3. Channel Mapping Strategy

Given the heterogeneity in channel counts (from 4 to 128), we employed a topological mapping strategy $\mathcal{R}(\cdot)$ mentioned in Section 3.2.

- **Sparse Electrodes:** We map channels based on their angular position around the forearm. For instance, Channel 1 is aligned to the *Palmaris Longus* muscle, and subsequent channels are ordered clockwise.
- **High-Density Arrays:** We apply average pooling over local regions to simulate virtual sparse channels when training the base model, or treat them as independent tokens with specific spatial embeddings in the Large model.
- **Normalization:** All signals are resampled to 200Hz to match the lowest common denominator for the base vocabulary training, ensuring temporal consistency.

C. Implementation Details

C.1. Experimental Setup

Pre-training Configuration Prior to the VQ neural vocabulary training(Section3.4) and the masked modeling pre-training(Section3.5) stages, all EMG recordings in the aggregated dataset are first downsampled to a uniform frequency. Subsequently, they are converted into a unified EMG sentence format using the neuromuscular contraction tokenizer defined in Section3.2.

The dataset is partitioned into training and validation sets at a 9:1 ratio. The validation set data are utilized solely for loss calculation during the VQ training phase.

- **VQ Vocabulary Training Stage** The learning rate is set to 5×10^{-5}, and the model is trained for 100 epochs. Following the method described in Section3.4, the EMG vocabulary size (k) is set to 8192, and the dimension (d) of the collective muscle contraction tokens is set to 256.
- **Masked Modeling Pre-training Stage:** The learning rate is set to 5×10^{-4}, and the training lasts for 200 epochs. A random masking ratio of 50% is applied.

In both aforementioned stages, a randomly initialized NST is employed as the backbone network. Adam optimizer is selected, and a cosine annealing schedule is applied for learning rate adjustment. All pre-training experiments are conducted on 8 NVIDIA GeForce RTX 4090 GPUs, with a batch size of 64 per GPU. A global random seed is set to 0 to ensure the reproducibility of all results.

Downstream Task Configuration The proposed method is evaluated on the four widely used public datasets listed below, which cover dozens of gestures. Detailed information regarding the data splits can be found in the Appendix. These datasets encompass diverse gesture sets and electrode configurations:

- **Dataset 1 (ULB-MLG) [13]:** Contains 28 subjects (14 participants, with each hand treated as an independent subject) performing 4 gestures plus a rest state, with each gesture repeated 30 times.
- **Dataset2 (EMG-EPN-612) [3]:** A large-scale dataset. Data from the first 100 subjects are used, containing 5 gestures plus a rest state, repeated 25 times each.
- **Dataset3 (NinaproDB4) [12]:** Includes 10 subjects and evaluates 10 selected gestures.
- **Dataset4 (Toro-Ossaba) [14]:** Comprises 8 participants performing 5 hand postures, with each posture held for 20 seconds.

Prior to downstream task fine-tuning, all recordings from the evaluation datasets are first downsampled to 200 Hz, consistent with the pre-training frequency, and then converted into uniform EMG sentences using the NCT.

For the evaluation protocol, the most stringent LOSO-CV [15] is adopted. In the linear probing evaluation, the parameters of the pre-trained NST backbone are kept frozen. Only a randomly initialized linear classification head is trained. The learning rate is set to 5×10^{-3} and training proceeds for 100 epochs. Accuracy serves as the core evaluation metric for all downstream tasks. The random seed is consistently set to 0 to ensure reproducibility.

D. Additional Experiments and Analysis

D.1. Ablation on Codebook Size

The vocabulary size k serves as a fundamental hyperparameter in the NCT, determining the semantic granularity of the learned physiological language. An appropriate vocabulary size must strike a critical balance between representational capacity and generalization efficiency. To systematically investigate this trade-off, a series of experiments are conducted using the Ninapro DB4 dataset. The codebook size k is varied across the set $\{1024, 2048, 4096, 8192, 16384\}$, while the embedding dimension d is fixed at 256 and consistent training protocols are maintained for the VQ process.

The quantitative results are presented in Table 3. A distinct trend is observed where downstream classification accuracy improves as the vocabulary size increases, reaching an optimal peak at $K = 8192$. This trajectory suggests that the model requires a sufficiently rich vocabulary to encode the complex spatiotemporal dynamics of EMG signals. Detailed analysis reveals distinct phenomena at the extremes of the spectrum:

- **Under-segmentation (Small k):** With a restricted vocabulary (e.g., $k = 1024$), the model suffers from morphological under-fitting. In this regime, the VQ codebook is forced to compress distinct muscle activation patterns into shared discrete tokens. This semantic blurring results in the loss of fine-grained motor details, making it difficult for the model to distinguish between subtle gesture variations, thereby capping the classification accuracy at 81.42%.
- **Optimal Granularity ($k = 8192$):** The performance maximization at $k = 8192$ indicates that this size provides sufficient capacity to cover the diverse range of motion primitives present in the multi-subject dataset without introducing excessive noise. This finding aligns with the complexity of human motor control, suggesting that a vocabulary of this scale effectively captures the fundamental phonemes of the myoelectric language.
- **Over-segmentation (Large k):** When the vocabulary is expanded beyond the optimum to 16384, a slight performance degradation is observed (87.85%). This decline is attributed to feature sparsity and semantic redundancy. An excessively large codebook tends to oversegment physiologically similar patterns into disjoint tokens, creating synonyms that the model fails to equate. Furthermore, the long-tail tokens in a large vocabulary may encode subject-specific noise or artifacts rather than generalizable muscle semantics, increasing the difficulty of the masked modeling pretext task and hindering cross-subject transfer.

Table 3. Ablation study on Vocabulary Size (k) on Ninapro DB4 (LOSO Accuracy).

Vocabulary Size (k)	**Accuracy (%)**
1024	81.42
2048	84.90
4096	86.15
8192 (Default)	**88.10**
16384	87.85

D.2. Ablation on Mask Ratio

In the MEM pre-training stage, the mask ratio is a critical hyperparameter. We evaluated mask ratios of $\{0.15, 0.30, 0.50, 0.75\}$. Table 4 indicates that a high mask ratio ($50\% \sim 75\%$) is beneficial for EMG. Unlike NLP typically 15% or Image typically 75% in MAE, EMG signals have high temporal redundancy. A higher mask ratio forces the model to learn long-range temporal dependencies rather than relying on local interpolation.

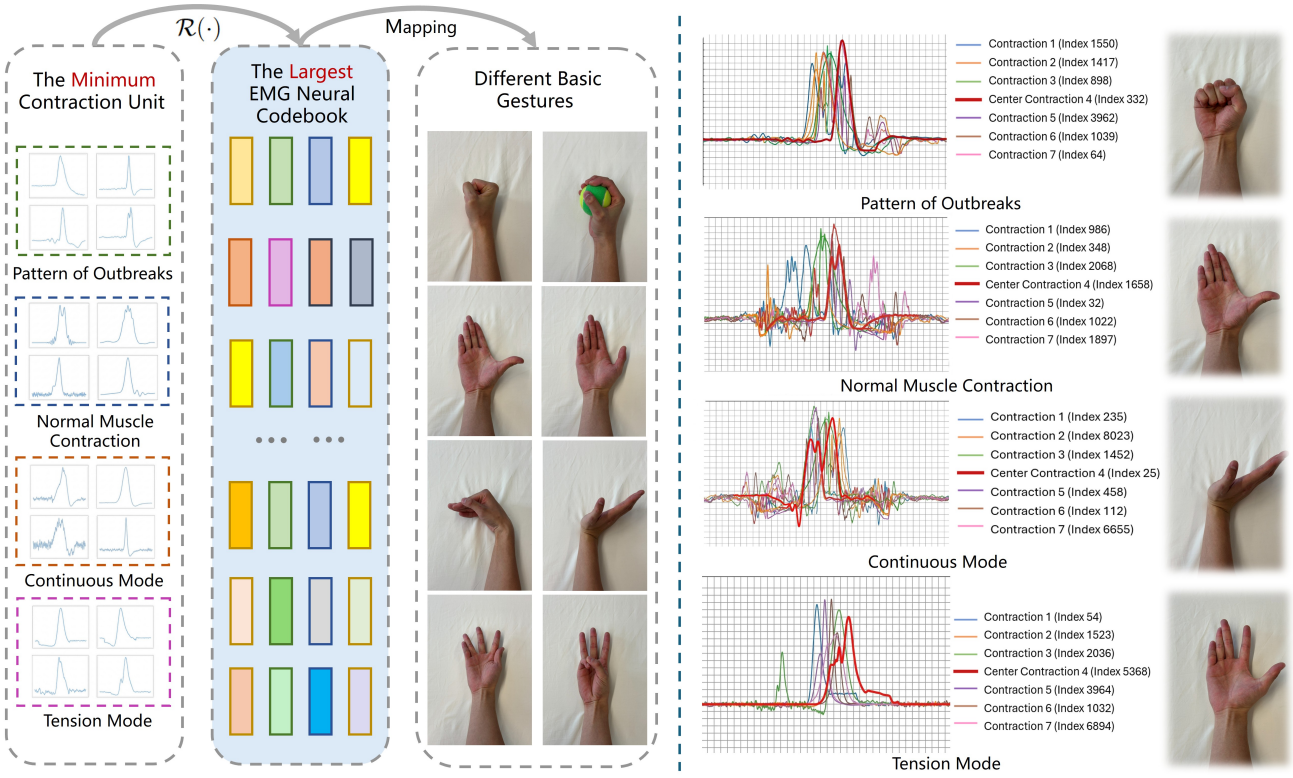

Figure 4. EMG vocabulary visualization

Table 4. Impact of Mask Ratio on downstream performance.

Mask Ratio	Accuracy (%)
15%	82.30
30%	85.12
50% (Default)	**88.10**
75%	87.65

D.3. Robustness Analysis

A critical imperative in deploying surface EMG-based interfaces is ensuring that the learned representations capture robust motor intentions rather than overfitting to device-specific artifacts or environmental interference. To rigorously evaluate the noise resilience of the proposed framework, an increasing level of Additive White Gaussian Noise (AWGN) is systematically injected into the test samples during the inference phase. The Signal-to-Noise Ratio (SNR) is varied across $\{20\text{dB}, 10\text{dB}, 5\text{dB}\}$ to simulate environments ranging from mild disturbances to extreme signal contamination.

The quantitative comparisons are detailed in Table 5. Under clean conditions, AEMG already establishes a performance lead. However, the disparity becomes significantly more pronounced as the noise intensity increases. Specifically, in the most severe scenario (SNR $=$ 5dB), the AEMG framework demonstrates remarkable stability, maintaining a high accuracy of 80.2%. In sharp contrast, the baseline models, ResNet-1D and the Transformer trained from scratch, exhibit precipitous performance degradation, plummeting to 55.4% and 59.1%, respectively. While the baselines suffer a relative performance drop of nearly 30% under extreme noise, the AEMG model incurs a minimal loss, evidencing its superior stability.

This resilience stems from the EMG-as-a-Language paradigm. Conventional end-to-end approaches directly map continuous raw signals to labels, rendering them susceptible to high-frequency noise perturbations. In contrast, AEMG processes discretized muscle contraction tokens. The vector quantization stage inherently functions as a nonlinear filter that eliminates low-amplitude noise while preserving dominant semantic structures of muscle activation. Furthermore, the masked EMG modeling pre-training compels the model to utilize global contextual dependencies rather than local signal morphology, thereby preventing overfitting to superficial noise patterns.

D.4. Codebook Utilization Analysis

To verify the validity of the EMG-as-a-Language metaphor and ensure no mode collapse occurred, we analyzed the us-

Table 5. **Robustness comparison.** AEMG shows minimal performance drop under severe noise.

Method	Clean	5dB	10dB	20dB
ResNet-1D (Baseline)	82.2	77.1	67.0	55.4
Transformer (Scratch)	83.5	79.3	71.8	59.1
AEMG (Ours)	**88.1**	**86.8**	**84.7**	**80.2**

Table 6. Ablation study results for different embeddings in the NST architecture.

Method	ULB-MLG[13]	Ninapro DB4[12]
AEMG-Large (Ours)	**91.50**	**88.10**
w/o Spatio-temporal Embedding	87.40	84.15
w/o Spatial Embedding	88.70	85.30
w/o Temporal Embedding	89.70	86.00

age frequency of the codebook vectors on the test set. Results indicate that with a codebook size of $k = 2048$, **96.8%** of the tokens were activated at least once. The distribution of token usage follows a long-tail but non-degenerate distribution, confirming that the NCT effectively maps continuous signals to a rich discrete semantic space.

D.5. Ablation Experiments for Various Embeddings of NST

An ablation study is conducted on the spatio-temporal embedding mechanism in NST (Section 3.3). As shown in Table 6, the experiments are performed on ULB-MLG and Ninapro DB4 with leave-one-subject-out accuracy reported. Removing spatial embedding Spatial Embedding prevents the model from distinguishing tokens from different physical channels, leading to a notable performance decrease of 2.8% on Dataset 3, where spatial information is critical. Removing temporal embedding Temporal Embedding causes the loss of timing information for contraction events, resulting in performance degradation on both datasets. The most severe decline occurs when both embeddings are removed Spatio-temporal Embedding, with a 4.1% drop on Dataset 1. These findings confirm the effectiveness of NST's design, demonstrating that integrating spatio-temporal information significantly enhances EMG signal interpretation and improves downstream task performance.

D.6. Mask modeling pretraining across patches

The pre-training process of the AEMG framework is illustrated in Step 2 of the fig.2 in the main paper. During this stage, the framework learns generalized spatiotemporal-level and synergistic-level muscle activation representations by masking a subset of individual muscle contraction tokens in EMG sentences and predicting the corresponding collective muscle contraction tokens at the masked positions based on the unmasked context.

D.6.1. Masking of individual muscle contraction words

To enable the AEMG framework to learn features of EMG sentences, including temporal dynamics and synergistic patterns of muscle contractions, a random masking operation is applied to individual muscle contraction tokens. This design forces the model to comprehend the contextual information of the entire EMG sentence to accurately reconstruct the masked portions.

For an EMG sentence $\mathbf{I} \in \mathbb{R}^{L\times T}$ obtained through NCT, its individual tokens are denoted as $\mathbf{e} = \{e_i \mid i = 1, \ldots, l\}$. A random binary mask $\mathbf{M} = \{m_i \mid i = 1, \ldots, l\}$ with $m_i \in \{0, 1\}$ is generated [6]. The masked EMG sentence $\mathbf{I}_M$ is then constructed by replacing masked tokens with a learnable mask token $e_{\text{mask}} \in \mathbb{R}^T$ [7]:

$$\begin{aligned}\mathbf{I}_M = &\{e_i \text{ if } m_i = 0 \mid i = 1, \ldots, l\} \\ &\cup \{e_{\text{mask}} \text{ if } m_i = 1 \mid i = 1, \ldots, l\}\end{aligned} \tag{4}$$

D.6.2. Prediction of collective muscle contraction words

This stage aims to predict the indices of collective muscle contraction tokens at masked positions (where e_{mask} is placed) based on the unmasked contextual tokens. The objective is achieved by minimizing the discrepancy between predicted and ground-truth token indices (obtained via vector quantization).

The masked sentence $\mathbf{I}_M$ is fed into the ST-EMGFormer to extract hidden vectors $\mathbf{h} = \{h_i \mid i = 1, \ldots, l\}$. These hidden representations are then projected through a linear classifier to predict the corresponding collective tokens:

$$p(v' \mid e_{\text{M}}) = \text{softmax}(\text{Linear}(h_i)) \tag{5}$$

where v' denotes any collective token in the vocabulary $\mathcal{V}$ [7]. The training objective for this MEM stage is defined as [6]:

$$\mathcal{L}_{\text{MEM}} = -\sum_{\mathbf{I}\in\mathcal{D}} \sum_{m_i=1} \log p(v_{z_i} \mid e_{\text{M}}) \tag{6}$$

D.7. Visualization of Neuromuscular Actions

To elucidate the physiological underpinnings of the learned representations, we analyze the characteristics of the constituent Motor Unit Action Potentials (MUAPs) across different movement primitives[8, 10]. Specifically, it provides a comprehensive spatiotemporal profile of the extracted MUAPs. The multi-channel waveforms reveal unique, movement-specific temporal signatures; the individual traces (thin lines) cluster tightly around the median waveform (thick lines), demonstrating the high consistency of the neural drive for actions such as Index Flexion (IF) and

Wrist Pronation (WP). Complementing this temporal analysis, the spatial profiles map the distinct topographic distribution of muscle fiber activation around the forearm cross-section. The radar plots illustrate that different movements activate specific angular sectors corresponding to underlying muscle compartments, creating a unique spatial fingerprint for each gesture. Collectively, these analyses confirm that the EMG signals encoded by our framework are grounded in consistent, physiologically distinct motor unit activities, effectively serving as the fundamental phonemes of the neuromuscular language.

E. Societal Impact and Limitations

Societal Impact: AEMG paves the way for plug-and-play prosthetic control and accessible HCI, potentially improving the quality of life for amputees. However, the decoding of motor intent raises privacy concerns regarding biometric data, which should be addressed with secure on-device processing.

Limitations: Currently, AEMG relies on a fixed channel mapping for topology alignment. Future work could explore graph-based approaches to handle arbitrary electrode placements dynamically without predefined mapping rules.

References

[1] Manfredo Atzori, Arjan Gijsberts, Simone Heynen, Anne-Gabrielle Mittaz Hager, Olivier Deriaz, Patrick Van Der Smagt, Claudio Castellini, Barbara Caputo, and Henning Müller. Building the ninapro database: A resource for the biorobotics community. In *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 1258–1265. IEEE, 2012. 1

[2] Manfredo Atzori, Arjan Gijsberts, Claudio Castellini, Barbara Caputo, Anne-Gabrielle Mittaz Hager, Simone Elsig, Giorgio Giatsidis, Franco Bassetto, and Henning Müller. Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Scientific Data*, 1(1):1–13, 2014. 2

[3] M.E. Benalcazar, L. Barona, L. Valdivieso, X. Aguas, and J. Zea. Emg-epn-612 dataset, 2020. Zenodo. 5

[4] Ulysse Côté-Allard, Cheikh Latyr Fall, Alexandre Drouin, Alexandre Campeau-Lecours, Clément Gosselin, Kyrre Glette, François Laviolette, and Benoit Gosselin. Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(4):760–771, 2019. 2

[5] Weidong Geng, Yu Du, Wenguang Jin, Wentao Wei, Yu Hu, and Jiajun Li. Gesture recognition by instantaneous surface emg images. *Scientific Reports*, 6(1):36571, 2016. 1, 2

[6] Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous EEG data in BCI. In *The Twelfth International Conference on Learning Representations*, 2024. 7

[7] J. Jin, H. Wang, H. Li, J. Li, J. Pan, and S. Hong. Reading your heart: Learning ecg words and sentences via pre-training ecg language model. *arXiv preprint arXiv:2502.10707*, 2025. 7

[8] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and A. J. Hudspeth. *Principles of Neural Science, Fifth Edition*. McGraw-Hill Medical, 2000. 1, 7

[9] N. Krilova, I. Kastalskiy, V. Kazantsev, V. A. Makarov, and S. Lobov. Emg data for gestures. UCI Machine Learning Repository, 2019. DOI: https://doi.org/10.24432/C5ZP5C. 2

[10] R. Merletti and D. Farina. *Surface Electromyography: Physiology, Engineering, and Applications*. John Wiley & Sons, 2016. 1, 7

[11] Mehmet Akif Ozdemir, Deniz Hande Kisa, Onan Guren, and Aydin Akan. Dataset for multichannel surface electromyography (semg) signals of hand gestures. *Data in Brief*, 41: 107921, 2022. 2

[12] Stefano Pizzolato, Luca Tagliapietra, Matteo Cognolato, Monica Reggiani, Henning Müller, and Manfredo Atzori. Comparison of six electromyography acquisition setups on hand movement classification tasks. *PLoS ONE*, 12(10): e0186132, 2017. 2, 5, 7

[13] C. Simar, M. Colot, A.-M. Cebolla, M. Petieau, G. Cheron, and G. Bontempi. Machine learning for hand pose classification from phasic and tonic emg signals during bimanual activities in virtual reality. *Front. Neurosci.*, 18:1329411, 2024. 5, 7

[14] A. Toro-Ossaba, J. Jaramillo-Tigreros, J.C. Tejada, A. Peña, A. López-González, and R.A. Castanho. Lstm recurrent neural network for hand gesture recognition using emg signals. *Appl. Sci.*, 12(19):9700, 2022. 5

[15] J. Yang, M. Soh, D. J. Weber, V. Lieu, and Z. Erickson. Emgbench: Benchmarking out-of-distribution generalization and adaptation for electromyography. *arXiv preprint*, 2024. arXiv:2410.23625. 1, 2, 5

[16] Liangtao Yang, Lu Gan, Zhenggang Zhang, Zhilin Zhang, Hui Yang, Yi Zhang, and Jinglong Wu. Insight into the contact impedance between the electrode and the skin surface for electrophysical recordings. *ACS omega*, 7(16):13906–13912, 2022. 1