ReactEMG: Zero-Shot, Low-Latency Intent Detection via sEMG

Runsheng Wang* **Xinyue Zhu*** **Ava Chen** **Jingxi Xu** **Lauren Winterbottom**
Dawn M. Nilsen **Joel Stein** **Matei Ciocarlie**

Columbia University

Abstract

Surface electromyography (sEMG) signals show promise for effective human–computer interfaces, particularly in rehabilitation and prosthetics. However, challenges remain in developing systems that respond quickly and reliably to user intent, across different subjects and without requiring time-consuming calibration. In this work, we propose a framework for EMG-based intent detection that addresses these challenges. Unlike traditional gesture recognition models that wait until a gesture is completed before classifying it, our approach uses a segmentation strategy to assign intent labels at every timestep as the gesture unfolds. We introduce a novel masked modeling strategy that aligns muscle activations with their corresponding user intents, enabling rapid onset detection and stable tracking of ongoing gestures. In evaluations against baseline methods, considering both accuracy and stability for device control, our approach surpasses state-of-the-art performance in zero-shot transfer conditions, demonstrating its potential for wearable robotics and next-generation prosthetic systems. Our project page is available at: `https://reactemg.github.io/`

1 Introduction

Learning-based methods are seeing increasing use for human-machine interfaces, where they enable users to control a wide range of robotic systems through interactive modalities such as wearable sensors, extended reality, and voice commands. Among these modalities, surface electromyography (sEMG) stands out for its non-invasive ability to detect muscle activity directly at the skin's surface. Direct sEMG measurement of neuromuscular signals has enabled non-clinical uses that include commanding multi-fingered grippers [1, 2], humanoid robots [3], and drones [4], as well as clinical applications, such as diagnosing neuromuscular conditions and controlling prosthetic limbs [5, 6].

Despite these promising advances, a number of important challenges remain for EMG signal interpretation. These are particularly apparent when human gestures (or intended gestures) derived from EMG signals are used for operating wearable manipulators, such as hand prostheses, exoskeletons or teleoperation systems. Human operators exhibit limited patience for perceived delay between a user input and robot execution—wait too long and the human will attempt different control inputs to compensate [7–9], which can lead to unstable robot behaviors [10, 11]. In the rehabilitation domain, building continuous, strong associations between human exertion and observable movement is key for achieving therapeutic outcomes [12–14], and executing robotic movement under low-latency constraints is thus necessary for closing the human-motion loop to give therapeutic benefits [15, 16].

To meet the constraints required by these application classes, we focus here on methods for gesture classification from EMG data that exhibit the following characteristics, illustrated in Figure 1:

*Equal contribution.

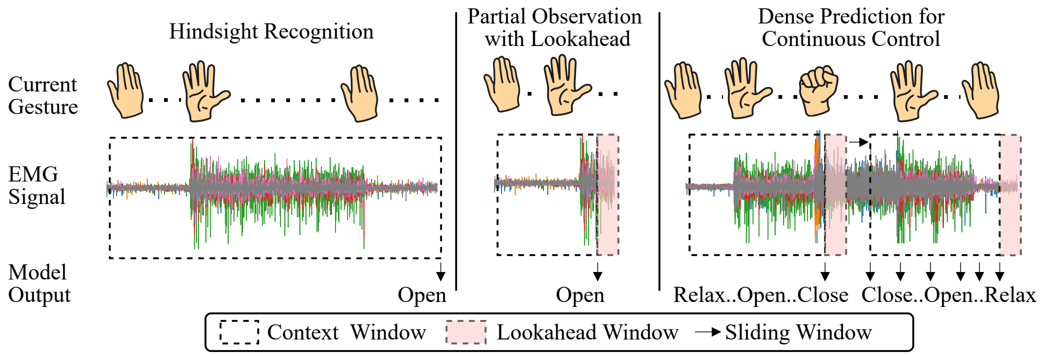

Figure 1: Paradigms for EMG-based intent detection. **Left:** a gesture recognition method that assumes full observability from gesture onset to offset, and assigns a single label retrospectively to the entire gesture. Such methods are not applicable to continuous device control. **Middle:** an intent detection method that assigns a label to a specific timestep, using a "lookahead" window of data recorded after the timestep being labeled. Limiting the span of the lookahead window allows intent detection with low latency. **Right:** the same method as in the middle, used in a sliding window paradigm to label a continuous stream of incoming EMG data, potentially containing successive transitions between different gestures as well as stable maintenance periods.

- **Dense labeling**: unlike gesture recognition methods that consume input data corresponding to one single gesture from onset to offset, then classify it in hindsight [17–19] (Fig. 1 Left), intent detection methods for device control must continuously interpret partial observations and output labels *as gestures unfold*, and recognize consecutive transitions between multiple different gestures.
- **Limited lookahead for low latency**: consider an intent detection model that, in order to classify the intent at time t, uses EMG data collected up to time $t + \ell$ (Fig. 1 Middle). Using $\ell > 0$ can be beneficial for the model, providing additional data as the new gesture unfolds. However, since the prediction must now lag behind ground truth intent change by a time interval of at least length ℓ, using large values for ℓ introduces latency in the response. We thus focus on methods where the value of ℓ, dubbed here "lookahead" time, is low enough to ensure responsiveness.
- **Stability**: while the ideal model must recognize successive intent changes, it must also deal appropriately with long "maintenance windows" where a single intent is maintained for longer time periods. During such windows, inadvertent "flicker" in the detected intent can have negative consequences if it produces corresponding flicker in the device being controlled. We thus focus on methods that distinguish between transient activity at a gesture's onset and sustained activation throughout its maintenance (Fig. 1 Right), and do not predict false transitions during the latter.

These desired characteristics are partially captured by the traditional metric of per-time-step classification accuracy, dubbed here "raw accuracy". While good performance on raw accuracy is important, it can also hide situations where a transition to a new gesture is detected with delay, or a maintenance window exhibits unwanted flicker. To capture such cases, we use a new metric dubbed "transition detection accuracy". According to this metric, a transition is considered to be successfully detected only if the model output correctly transitions between intents close enough to the ground truth change, and exhibits no instability either before or after the transition.

To achieve high performance on both these metrics, this paper presents a combination of a novel learning model, training framework, dataset, and evaluation for low-latency intent detection via EMG. We treat EMG intent detection as a segmentation problem, predicting an action label (i.e. user intent) at each timestep within a continuous EMG window. To train, we provide the model with both EMG signals and corresponding actions as input, then adopt a masked modeling strategy that selectively masks portions of these sequences. By requiring the model to reconstruct the missing segments, we leverage local supervision that anchors muscle activations to the subject's intent. This approach enables the model to learn robust signal–action alignments, even under imprecise ground truth labels. In summary, our contributions are as follows:

- **Masked modeling for low-latency EMG segmentation**: We propose a novel approach consisting of a masked modeling-based segmentation architecture for intent detection via EMG. The model jointly learns EMG and user intent, continuously making predictions at every timestep, end-to-end, without any manual feature engineering and thresholding. By aligning EMG signals to user actions through local supervision, our approach captures both transient onset patterns and sustained activations for low-latency intent detection. In addition, we incorporate unlabeled EMG recordings, allowing the model to discover intrinsic muscle synergies without explicit labels. By learning from both labeled and unlabeled data, the model becomes more robust to anatomical differences, electrode placement shifts, and physiological variations, ultimately improving its generalization.
- **Existing and novel data:** we train on a collection of open-source EMG datasets spanning diverse participants and tasks. We also collect and release a new dataset capturing various arm positions and grasping motions. Our dataset comprises a wider variety of functional situations, such as transitions between multiple different gestures, maintenance windows of varying length, etc.
- **Zero-shot generalization:** leveraging a diverse corpus of labeled and unlabeled EMG data—along with our newly collected dataset—our model maintains high performance on both metrics of interest without any subject-specific calibration. Gathering fine-tuning EMG data through repetitive gestures can be time-consuming, physically demanding, and logistically complex, especially for people with limited mobility. However, developing a zero-shot model remains challenging because EMG signals can differ widely from one person to another, reflecting anatomical differences, physiological factors, and session-to-session variability.
- **Accuracy:** Our method surpasses the state of the art on the traditional metric used to quantify intent detection performance, namely per-time step dense labeling (raw accuracy). We also propose a new metric, which jointly considers how quickly the system reacts to transitions and how reliably it maintains commands (transition accuracy). Our method also surpasses baselines in this highly demanding test, suggesting promising future applications to device control.

2 Related work

Gesture recognition via sEMG has long been investigated as a control signal for human-computer interaction (HCI), particularly for prosthetic devices, assistive robotics, and rehabilitation systems [20–27]. Earlier approaches predominantly relied on extracting handcrafted features from the sEMG signals followed by traditional machine learning classifiers such as Linear Discriminant Analysis [28] or Support Vector Machines [29–31]. Recent research has shifted focus to deep learning paradigms, leveraging diverse network architectures and signal representations. Betthauser et al. showed that sequential models, including temporal convolutional networks, significantly improve prediction stability and latency compared to static-feature classifiers [17]. Along similar lines, Simão et al. deployed recurrent neural networks for online EMG gesture classification [18], and subsequent LSTM-based methods demonstrated that modeling temporal dependencies yields higher accuracy than conventional feature-driven approaches [19, 32].

Beyond core intent detection tasks, several large-scale sEMG benchmarks have been released to investigate broader challenges in muscle-driven interfaces. Sivakumar et al. introduce a dataset of wrist sEMG from 108 users paired with keystroke transcripts for speech-recognition–style sequence modeling, but observe a sharp performance drop when generalizing to unseen users without adaptation [33]. Salter et al. introduces a dataset of 16-channel HD-sEMG synchronized to 26-camera motion capture across 193 users and 29 gesture stages [34]. Their pipeline excels at reconstructing continuous joint kinematics, though it does not emphasize higher-level semantic intent recognition, which need to accommodate the natural variations in how different individuals perform the same gesture. Meanwhile, Yang et al. aggregate nine public gesture corpora to benchmark out-of-distribution classification via leave-one-subject-out and few-shot adaptation, but do not address streaming, low-latency decoding or rapid calibration-free deployment [35].

Transformer architectures have also emerged as a compelling approach for capturing the complex spatiotemporal dynamics inherent to sEMG. Montazerin et al. proposed a Vision Transformer framework (CT-HGR) for amputee hand gesture recognition using 128-channel HD-sEMG inputs, achieving high accuracy even with short observation windows [36]. Zabihi et al. introduced a hybrid transformer–convolution architecture (TraHGR), highlighting the advantage of attention mechanisms for modeling distributed muscle activation patterns, though their approach remains fully supervised [37]. In addition, reinforcement learning has been explored by Cruz et al., who framed

EMG classification as a sequential decision-making task, training a Deep Q-Network agent to classify windowed EMG and IMU signals [38].

Most relevant to our work are methods targeting user-independent or calibration-free EMG intent recognition. Transfer learning and domain adaptation have shown promise in handling the intrinsic variability and non-stationarity of sEMG signals [39–43]. Xu et al. proposed an autoregressive generative model that produces synthetic sEMG streams conditioned on a small "prompt" snippet from a new user or session [44], greatly reducing per-user data requirements. Likewise, La Rotta et al. introduced a meta-learning framework that rapidly fine-tunes to new users based on experience drawn from many subjects [45]. Finally, Schiel et al. tackled signal drift and session variability with an incremental learning approach using sparse Gaussian Processes, continuously updating the classifier in an unsupervised manner to adapt to changing conditions [46].

Most existing methods focus on gesture recognition with retrospective classification, leading to delays and a lack of real-time performance. Our method, however, provides continuous intent detection with low-latency predictions at every timestep. We introduce a masked modeling strategy that aligns EMG signals with user intent, enabling faster and more accurate predictions. Unlike existing approaches that require extensive calibration for each user, our model generalizes to new subjects without additional training. Additionally, while other methods suffer from instability or "flickering" during sustained gestures, our approach maintains stable intent detection over long periods. These differences make our method more efficient and reliable for real-world applications.

3 Method

Our goal is to continuously predict user intent from forearm muscle activity. We capture an 8-channel EMG signal (Myo armband, Thalmic Labs), sampled at 200 Hz, and train a model to predict the current intent from a predefined set of actions at every instant. We formulate this as a segmentation problem over a continuous stream of muscle activity. Let $X = \{x_1, x_2, \dots, x_T\}$ represent a window of length T of multichannel EMG time series, where $x_t \in \mathbb{R}^C$ corresponds to the EMG measurements from C sensors (e.g., eight Myo armband channels) at time t. We define a discrete set of actions $\mathcal{A} = \{a_1, a_2, \dots, a_K\}$, such as "open hand" or "close hand." The goal is to learn a function $f : \mathbb{R}^{C\times T} \to \mathcal{A}^T$ that assigns, for every time index t, an action label $y_t \in \mathcal{A}$.

3.1 Architecture

Although segmentation captures continuous changes in muscle activation, the EMG signal itself remains non-stationary and highly variable across time. Even within a single recording session, different gestures can induce distinct activation patterns, and these patterns may fluctuate due to electrode shift, fatigue, or user variability. Inspired by recent advances in multimodal LLMs and "any-to-any" frameworks [47–49], we hypothesize that explicitly conditioning EMG representations on the corresponding action label—allowing the model to cross-attend between the two—can mitigate these challenges. To this extent, we propose an encoder-only Transformer for EMG and action modeling. Unlike conventional methods that treat EMG signals and their corresponding action labels as an input–output pair, our model treats EMG and intent as two distinct input modalities. We construct a multimodal sequence that includes EMG data followed by the corresponding intent data. The model's objective is to reconstruct the original unmasked sequence—both EMG and intent—at all timesteps. Figure 2 provides an overview of the proposed architecture.

Given a raw EMG signal, we first apply a median filter, then rectify negative amplitudes, and extract overlapping sliding windows. Our model treats EMG signals and subject intent as two separate modalities, each mapped to its own space. For EMG, we use a learnable linear projection to map the raw 8-channel signal directly into the embedding space, without any manual feature extraction. In parallel, we represent the subject's intent in the embedding space via a standard lookup table (covering valid intents plus a dedicated mask token).

Next, we apply span masking to the projected EMG and intent embeddings. Following a protocol similar to RoBERTa's [50] dynamic masking, we randomly mask out contiguous timesteps in either or both modalities and regenerate these masks at each training epoch, thereby exposing different regions of the same sequence to the model over time. After masking, we add modality-specific encoding vectors and positional encodings to both the EMG and action tokens, ensuring each token

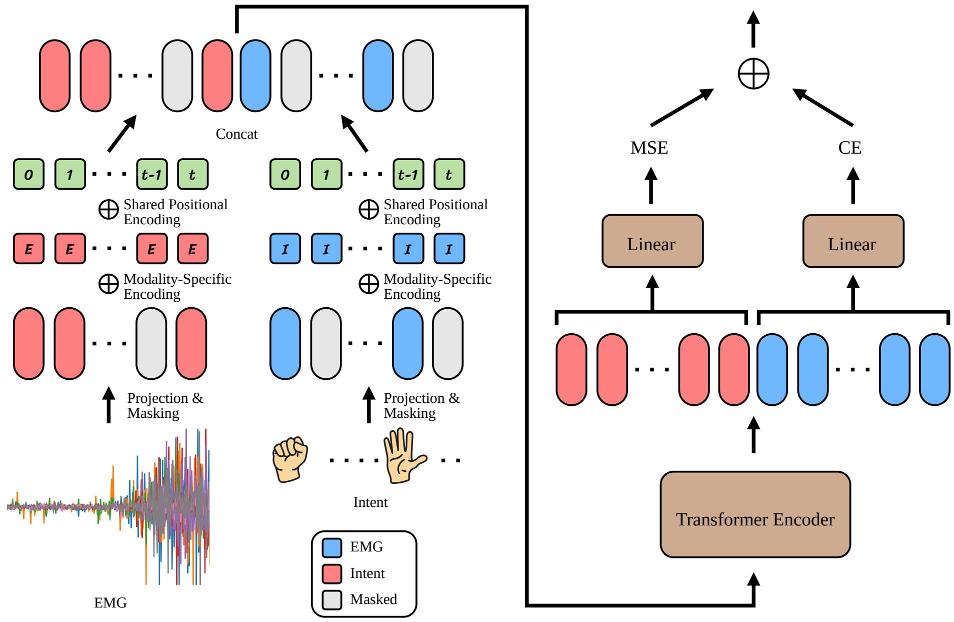

Figure 2: Model overview. EMG signals are mapped into an embedding space via a learnable linear projection, while intent tokens use a lookup-based embedding. Both modalities undergo dynamic span masking and receive modality-specific plus shared positional encodings. They are then concatenated and processed by a Transformer encoder, after which outputs are split into EMG and intent branches. The EMG branch is optimized via MSE on masked timesteps, and the intent branch via cross-entropy on masked tokens. Losses are added before backpropagation.

can be distinguished not only by its position in time but also by the modality it belongs to. These positional encodings are shared across both modalities to enforce consistent temporal alignment.

We then concatenate the EMG and intent sequences into a single multimodal sequence and feed it into a Transformer encoder to learn cross-modal dependencies. Following the Transformer encoder, we split the output multimodal sequence back into EMG and intent constituents. The EMG output is projected through a linear layer to produce eight values per timestep, and the loss on EMG is computed against raw 8-channel values using mean-squared error (MSE) over only the masked positions. The intent output goes through a separate linear layer for dense classification at every timestep, with a cross-entropy loss on masked intent tokens.

3.2 Masking strategy

Our approach centers on a multimodal masked reconstruction framework that unifies multiple training objectives by selectively masking portions of the EMG or intent sequences. The primary task is to reconstruct a fully masked intent sequence using only unmasked EMG, which aligns with the goal of intent detection. To help the model learn useful representations, we incorporate the following auxiliary tasks: **(1) Partial Intent Masking:** Only a subset of the intent tokens is masked, while the entire EMG signal and remaining intent context remain visible, helping the model learn to infer intent when full EMG data is present; **(2) Partial EMG Masking:** Only a subset of the EMG timesteps are masked but the intent tokens remain unmasked, prompting the model to reconstruct EMG from intent cues; **(3) Temporally Aligned Masking:** Both EMG and intent sequences are masked at the same timesteps, prompting the model to jointly recover the temporal relationship between muscle activations and intended actions.

Finally, we introduce a self-supervised EMG modeling task to exploit structure and synergies in unlabeled EMG data. Here, the model masks segments of the EMG signal while replacing intent tokens with the intent mask token and disabling attention to the intent sequence via an attention mask. This setup forces the model to reconstruct the missing EMG purely from its unmasked context. By integrating these tasks within one framework, the model not only learns to predict intent accurately but also develops a deeper cross-modal understanding of EMG and intent.

3.3 Online Inference

During online deployment, our trained model processes a continuous stream of EMG windows and generates a label for each timestep in every window. While we could assign a label to a timestep as soon as it appears in the sliding window, transient noise or incomplete gestures often lead to flickering. Since the model produces dense predictions and the windows overlap, multiple predictions are available for the same timestep. We have the option of fusing these overlapping predictions into a single label per timestep through an aggregator that balances speed and stability. Here, we present a smoothing method that applies to all models with dense output.

Lookahead for smoother predictions. Suppose we want to predict the intent at timestep t. To do this, we introduce a small lookahead parameter ℓ, which allows us to collect logits for timesteps from t to $t+\ell$ across subsequent windows. We then aggregate the model outputs from all windows that cover the interval $[t,\ t+\ell]$ by averaging their logits. Finally, we obtain the predicted class label for timestep t by applying argmax to this merged distribution.

Inference frequency. EMG sensors often operate at a high sampling frequency, while human intent changes and the corresponding muscle contractions occur at a much lower frequency [16]. A low-latency model does not need to perform inference at the sampling rate for human perception to remain unaffected by delays. To enhance prediction stability and reduce the reliance on specialized hardware for wearable robotics, we produce predictions at a lower frequency than the sensor's sampling rate. Instead of updating the label with every new sample, we finalize a label every s timesteps, and holding it constant for the next s timesteps. As a result, once a label is computed at timestep t, it remains fixed until $t+s$. By combining future context ℓ with a slower update rate s, this approach minimizes rapid fluctuations caused by noise while maintaining the system's ability to quickly adapt to genuine gesture changes. A detailed illustration of the proposed smoothing method is provided in the Supplementary Materials.

3.4 Datasets and training

Existing datasets: capturing the diversity of electromyography (EMG) signals is essential for robust generalization. However, many publicly available EMG datasets are recorded under fixed conditions, such as a specific posture, and do not reflect real-world variability. To address this, we aggregate several open-source EMG datasets [51–54], each offering distinct recording conditions and subject populations, thereby encompassing a broad range of muscle activation patterns.

The EMG-EPN-612 dataset [51] is particularly notable for its large size and subject population diversity. We report zero-shot results on the six-class EMG-EPN-612 dataset to benchmark our method against existing works in the literature. For other datasets that contain more than six classes, we select only the subset corresponding to EMG-EPN-612's six gestures as labeled data, and treat the remaining gestures as unlabeled for our self-supervised EMG modeling task. We do not include other datasets as we specifically focus on open-source datasets with a single Myo armband at 200 Hz and containing gestures that match those in the EMG-EPN-612 dataset.

When evaluating performance on the six-class EMG-EPN-612, we train the model exclusively on this dataset. We split the dataset training portion into training and validation subsets, holding out the test set entirely. We train for 12 epochs with a linear learning rate warmup and linear decay. We retain the checkpoint with the lowest validation loss across all epochs.

New dataset: to further evaluate our model's performance under more diverse conditions, we also collect our own dataset, with IRB approval. This new dataset comprises three gestures - open, close, and relax - recorded in four different arm positions, and two functional grasping tasks. We concentrate on three functional gestures as they especially relevant for controlling robotic devices such as a hand exoskeleton or a parallel gripper. In addition, our dataset contains examples of transitions between different gestures (as opposed to only transitions between relax and another gesture) to mirror more realistic use conditions. For a complete overview of the public EMG datasets and our data collection protocol, please see the Supplementary Materials.

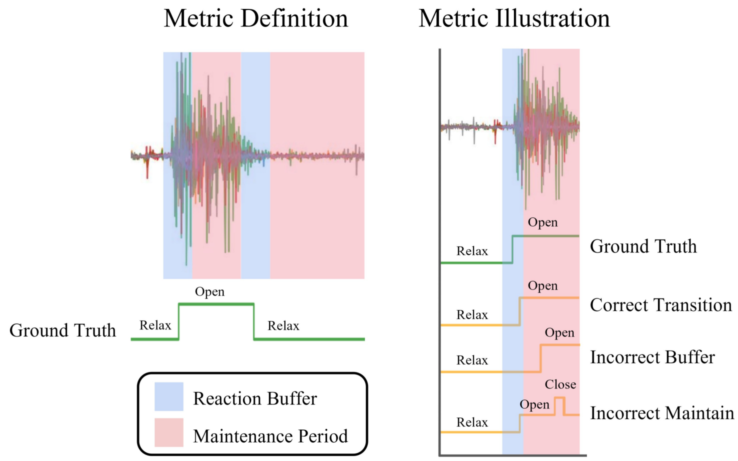

Figure 3: Illustration of the Transition Accuracy evaluation protocol. For each ground-truth change from class y_{old} to y_{new}, we define a *reaction buffer* (blue region) centered on the ground truth transition. The model must predict y_{new} at least once within this buffer and must not predict any other class. The subsequent *maintenance period* (red region) extends from the end of the reaction buffer to the start of the next reaction buffer; during this interval, the model's output must remain stably equal to y_{new}. A transition is scored as correct only if both reaction buffer and maintenance period conditions are satisfied.

When evaluating results on our new dataset, we adopt a pretraining / fine-tuning procedure. During pretraining, the model is trained on all available public datasets - labeled and unlabeled - using the same setup as in the EMG-EPN-612-only experiment. In the fine-tuning stage, the model is further trained on our dataset in a leave-one-subject-out manner: when testing on each subject, that subject's data is withheld from training. Fine-tuning runs for five epochs with no additional scheduler adjustments. Details of our architecture and training parameters are provided in the Supplementary Materials.

4 Evaluation

4.1 Metrics

Raw Accuracy: raw accuracy measures how often the predicted label at each EMG timestep matches the ground-truth label, serving as a straightforward proxy for overall performance in continuous control scenarios. Because our approach generates predictions at every EMG sample, it inherently supports per-timestep computations. For models that produce labels at coarser intervals, we upsample their outputs to the original EMG sampling rate to enable a fair, per-timestep comparison under realistic online conditions.

Transition Accuracy. Although raw accuracy captures per-timestep correctness, it does not account for how quickly the model responds to a new gesture or how consistently it maintains its prediction. To address these limitations, we propose Transition Accuracy, a more nuanced evaluation protocol that measures not only the onset of new gestures but also the stability of predictions over time.

Let the ground-truth label at time t be y_t. We define a "transition" as a segment of data that begins when the label transitions from a previous class y_{old} to a new class y_{new}, and ends at the next class switch. Each transition is divided into two disjoint intervals: **(1) Reaction buffer:** A short time window (typically 1 s in our experiments) centered on the ground truth intent transition timestep. The model must predict transition from y_{old} to y_{new} at least once within this buffer, and should not contain any other prediction classes. We use a buffer window in our transition accuracy metric to account for the fact that the exact timestamp of a transition is very difficult to identify even in ground truth data (since the time elapsed between an experimenter issuing an instruction and the subject executing the corresponding gesture can vary, and offline, manual labeling is subjective and impractical for large datasets). However, since we want models to be responsive in real-life use, we keep the buffer

Table 1: Comparison of Zero-Shot Average Accuracies on EMG-EPN-612 and Our Datasets.

		EPN 3-Class		EPN 6-Class		Our Dataset	
Method	**Lookahead**	**Raw Acc.**	**Transition Acc.**	**Raw Acc.**	**Transition Acc.**	**Raw Acc.**	**Transition Acc.**
ANN	0s	0.86	0.44	0.81	0.32	0.70	0.12
LSTM	0s	0.90	0.72	0.86	0.60	0.40	0.13
ED-TCN	0s	0.94	**0.79**	0.91	0.68	0.84	0.30
Ours	0s	**0.95**	**0.79**	**0.92**	**0.70**	**0.86**	**0.36**
LSTM	0.25s	0.92	0.77	0.88	0.66	0.82	0.33
ED-TCN	0.25s	0.94	0.79	0.91	0.68	0.86	0.41
Ours	0.25s	**0.95**	**0.82**	**0.92**	**0.74**	**0.89**	**0.52**

window short. **(2) Maintenance period**: The maintenance period spans the remainder of the gesture until the onset of the next reaction buffer. After predicting y_{new}, the model must consistently output the same label for the entire duration of this interval. Any incorrect prediction within this interval immediately invalidates the entire transition. Predictions on a transition is considered correct if and only if both these conditions are met. By penalizing both delayed onset detection and instability during the maintenance period, this metric provides a stringent and realistic measure of performance in real-world use cases.

4.2 Baselines

ED-TCN [17] is an encoder–decoder temporal convolutional network that relies solely on 1D convolutions. The encoder applies progressively downsampled temporal convolutions with large kernels, while the decoder mirrors this process through upsampling and additional convolutional layers. A final time-distributed dense layer then projects the decoder outputs into the target class dimension, producing per-segment gesture classifications.

LSTM [18] processes the input EMG signals in a sliding-window manner. Within each window, simple statistical features (e.g., standard deviations) are computed over short, overlapping sub-windows, forming a sequence of feature vectors. A fully connected layer first maps these vectors to a higher-dimensional space before feeding them into an LSTM layer that captures temporal dependencies. Finally, a dense layer provides gesture class scores at each timestep.

ANN [55] also employs a sliding window strategy but assigns each window a single label based on its final timestep. The preprocessing stage extracts a compact set of time-domain metrics—such as RMS, variance, and waveform length—from each channel in the window. After standardization, these features are fed into a multi-layer neural network, which outputs one predicted gesture label per window.

All baseline methods were carefully tuned to achieve strong performance on our task; detailed configurations and hyperparameters are provided in the Supplementary Materials.

4.3 Results and ablations

As shown in Table 1, our approach outperforms the state-of-the-art model on both metrics across all experimental setups. While the baseline methods experience a drop in performance when shifting from identifying gestures per-timestep (raw accuracy) to maintaining them over time (transition accuracy), our method remains more robust. This suggests that our masked modeling segmentation framework is particularly effective for continuous, low-latency EMG-based intent detection that requires rapid onset detection and stable gesture maintenance. We do not provide results for the ANN architecture with lookahead because its single-label nature precludes the use of smoothing.

We show results for two different lookahead window sizes: 0 s and 0.25 s. The size of the lookahead window is a lower bound on the latency introduced by the overall system: after the model has collected all data needed to label timestep t, including its corresponding lookahead window, it can then perform inference, after which it will output the label for timestep t. However, the inference time (typically less than 5 ms for our model) is small compared to the lookahead window, so we focus here on keeping the lookahead window short in order to decrease latency. We observe that a 0.25 s lookahead window helps performance (presumably by providing more context for the model), at a low cost in latency.

Table 2: Pretraining and Multi-Task Learning Ablations on Our Dataset.

Method	**Lookahead**	**Raw Acc.**	**Transition Acc.**
No Fine-tune	0s	0.61	0.02
No Pretrain	0s	0.68	0.03
No Multitask Learning	0s	0.85	0.34
Pretrain + Fine-tune	0s	**0.86**	**0.36**
No Fine-tune	0.25s	0.61	0.01
No Pretrain	0.25s	0.70	0.06
No Multitask Learning	0.25s	0.88	0.50
Pretrain + Finetune	0.25s	**0.89**	**0.52**

We also perform ablation studies to isolate the contributions of two key architectural design choices—pretraining and multi-task learning—summarized in Table 2. Additional ablation experiments, including the impact of context length and the effect of the reaction buffer size on the transition accuracy metric, are provided in the Supplementary Materials.

To assess the role of pretraining with public datasets, we compare three settings: (1) pretraining then fine-tuning, (2) training from scratch with no pretraining, and (3) pretraining only without finetuning. The pretraining-then-fine-tuning approach achieves the highest average intent detection accuracy, highlighting both the value of diverse open-source datasets collected on distinct subject population and the generalization benefits of large-scale pretraining. Additionally, when we evaluate the pretraining-only model directly on our dataset, its performance is poor—likely because the public datasets' recording conditions are small and fixed, whereas our dataset encompasses a broader range of conditions. We provide a more detailed analysis of these data distribution differences in the Supplementary Materials.

We also examine the impact of removing the auxiliary reconstruction objectives described in subsection 3.2. Training solely on intent reconstruction leads to a slight decrease in transition accuracy, indicating that multi-task reconstruction objectives could be beneficial for learning more robust features.

5 Conclusion

In this paper, we present a segmentation-based, zero-shot approach for low-latency sEMG intent detection. Our key insight is to jointly encode EMG signals and action labels in a Transformer architecture using a masked reconstruction framework, enabling the model to learn robust, fine-grained alignments between muscle activations and intended gestures. This design allows the system to infer user intent continuously—rather than waiting for gestures to complete—yielding faster onset detection and stable maintenance of ongoing commands. To characterize performance, in addition to the traditional, per-time-step raw accuracy metric, we introduce a new metric, dubbed transition accuracy, which evaluates how quickly the model reacts to changing muscle activity and how reliably it preserves each predicted intent. Across the EMG-EPN-612 dataset and our newly collected dataset, our method consistently outperforms state-of-the-art baselines on both metric, demonstrating its potential for low-latency, continuous applications such as wearable robotics, assistive devices, and prosthetic control.

6 Limitations

Our approach relies on a sparse, commercial eight-channel Myo armband positioned around the forearm, which concentrates on major extensor and flexor groups but may neglect smaller muscle groups and intrinsic hand muscles. Because of this placement, it can be challenging to distinguish between certain gestures that share overlapping muscle activations. For example, wrist extension naturally activates the same extensor group involved in opening the hand, leading to potential ambiguity. To better capture these nuanced differences, additional sensors or more sophisticated electrode arrangements would be necessary. Nonetheless, the Myo armband remains an accessible choice due to its commercial availability and the breadth of existing datasets, which facilitate large-scale training and comparative benchmarking.

Although our Transformer-based architecture improves on current state-of-the-art results, it requires substantially more computational resources than traditional models like SVMs or LDA. Even at a reduced inference frequency, preserving low latency necessitates running the model on a GPU, which may be impractical for wearable devices or other edge-computing scenarios. Techniques such as knowledge distillation could potentially alleviate these demands by creating lighter models suited to embedded devices. However, implementing these techniques introduces additional engineering complexities and trade-offs that are particularly challenging in resource-constrained environments.

References

[1] Aleksandra Loskutova, Daniel Roozbahani, Marjan Alizadeh, and Heikki Handroos. Design and development of a robust control platform for a 3-finger robotic gripper using EMG-derived hand muscle signals in NI LabVIEW. *Journal of Intelligent and Robotic Systems*, 110(3), September 2024. ISSN 1573-0409. doi: 10.1007/s10846-024-02160-w. URL `http://dx.doi.org/10.1007/s10846-024-02160-w`.

[2] Cassie Meeker, Maximilian Haas-Heger, and Matei Ciocarlie. A continuous teleoperation subspace with empirical and algorithmic mapping algorithms for nonanthropomorphic hands. *IEEE Transactions on Automation Science and Engineering*, 19(1):373–386, January 2022. ISSN 1558-3783. doi: 10.1109/tase.2020.3035156. URL `http://dx.doi.org/10.1109/TASE.2020.3035156`.

[3] Yunjun Nam, Bonkon Koo, Andrzej Cichocki, and Seungjin Choi. GOM-Face: GKP, EOG, and EMG-based multimodal interface with application to humanoid robot control. *IEEE Transactions on Biomedical Engineering*, 61(2):453–462, February 2014. ISSN 1558-2531. doi: 10.1109/tbme.2013.2280900. URL `http://dx.doi.org/10.1109/TBME.2013.2280900`.

[4] Yash Doshi and Divyanshi Nath. Designing a drone controller using electromyography signals. In *2021 International Conference on Communication information and Computing Technology (ICCICT)*, page 1–6. IEEE, June 2021. doi: 10.1109/iccict50803.2021.9510045. URL `http://dx.doi.org/10.1109/ICCICT50803.2021.9510045`.

[5] Jack Tchimino, Rehne Lessmann Hansen, Peter Holmberg Jørgensen, Jakob Dideriksen, and Strahinja Dosen. Application of EMG feedback for hand prosthesis control in high-level amputation: a case study. *Scientific Reports*, 14(1), December 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-80828-x. URL `http://dx.doi.org/10.1038/s41598-024-80828-x`.

[6] Jingxi Xu, Ava Chen, Lauren Winterbottom, Joaquin Palacios, Preethika Chivukula, Dawn M. Nilsen, Joel Stein, and Matei Ciocarlie. Reciprocal learning of intent inferral with augmented visual feedback for stroke, 2024. URL `https://arxiv.org/abs/2412.07956`.

[7] Sam Beech, Danaë Stanton Fraser, Andy Corston-Petrie, Andy P Gower, and Iain D Gilchrist. How changes in the mean latency, jitter amplitude, and jitter frequency impact target acquisition performance. *ACM Trans. Appl. Percept.*, 22(2):1–18, April 2025. URL `https://dx.doi.org/10.1145/3701984`.

[8] Federico Scholcover and Douglas J Gillan. Using temporal sensitivity to predict performance under latency in teleoperation. *Hum. Factors*, 60(1):80–91, February 2018. URL `https://dx.doi.org/10.1177/0018720817734727`.

[9] Jing Du, William Vann, Tianyu Zhou, Yang Ye, and Qi Zhu. Sensory manipulation as a countermeasure to robot teleoperation delays: system and evidence. *Sci. Rep.*, 14(1):4333, February 2024. URL `https://dx.doi.org/10.1038/s41598-024-54734-1`.

[10] Stefan Neumeier, Philipp Wintersberger, Anna-Katharina Frison, Armin Becher, Christian Facchi, and Andreas Riener. Teleoperation: The holy grail to solve problems of automated driving? sure, but latency matters. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, New York, NY, USA, September 2019. ACM. URL `https://dx.doi.org/10.1145/3342197.3344534`.

[11] Amro Khasawneh, Hunter Rogers, Jeffery Bertrand, Kapil Chalil Madathil, and Anand Gramopadhye. Human adaptation to latency in teleoperated multi-robot human-agent search and rescue teams. *Autom. Constr.*, 99:265–277, March 2019. URL `https://dx.doi.org/10.1016/j.autcon.2018.12.012`.

[12] Kristan A Leech, Ryan T Roemmich, James Gordon, Darcy S Reisman, and Kendra M Cherry-Allen. Updates in motor learning: Implications for physical therapist practice and education. *Phys. Ther.*, 102(1), January 2022. URL `https://dx.doi.org/10.1093/ptj/pzab250`.

[13] Silvestro Micera, Matteo Caleo, Carmelo Chisari, Friedhelm C Hummel, and Alessandra Pedrocchi. Advanced neurotechnologies for the restoration of motor function. *Neuron*, 105(4):604–620, February 2020. URL `https://dx.doi.org/10.1016/j.neuron.2020.01.039`.

[14] Neville Hogan, Hermano I Krebs, Brandon Rohrer, Jerome J Palazzolo, Laura Dipietro, Susan E Fasoli, Joel Stein, Richard Hughes, Walter R Frontera, Daniel Lynch, and Bruce T Volpe. Motions or muscles? some behavioral factors underlying robotic assistance of motor recovery. *J. Rehabil. Res. Dev.*, 43(5):605–618, August 2006. URL `https://dx.doi.org/10.1682/jrrd.2005.06.0103`.

[15] Ana Cisnal, Javier Perez-Turiel, Juan-Carlos Fraile, David Sierra, and Eusebio de la Fuente. RobHand: A hand exoskeleton with real-time EMG-driven embedded control. quantifying hand gesture recognition delays for bilateral rehabilitation. *IEEE Access*, 9:137809–137823, 2021. URL `https://dx.doi.org/10.1109/access.2021.3118281`.

[16] Andrew B Schwartz. Movement: How the brain communicates with the world. *Cell*, 164(6): 1122–1135, March 2016. URL `https://dx.doi.org/10.1016/j.cell.2016.02.038`.

[17] Joseph L. Betthauser, John T. Krall, Shain G. Bannowsky, Gyorgy Levay, Rahul R. Kaliki, Matthew S. Fifer, and Nitish V. Thakor. Stable responsive EMG sequence prediction and adaptive reinforcement with temporal convolutional networks. *IEEE Transactions on Biomedical Engineering*, 67(6):1707–1717, June 2020. ISSN 1558-2531. doi: 10.1109/tbme.2019.2943309. URL `http://dx.doi.org/10.1109/TBME.2019.2943309`.

[18] Miguel Simão, Pedro Neto, and Olivier Gibaru. EMG-based online classification of gestures with recurrent neural networks. *Pattern Recognition Letters*, 128:45–51, December 2019. ISSN 0167-8655. doi: 10.1016/j.patrec.2019.07.021. URL `http://dx.doi.org/10.1016/j.patrec.2019.07.021`.

[19] Lorena Isabel Barona López, Francis M. Ferri, Jonathan Zea, Ángel Leonardo Valdivieso Caraguay, and Marco E. Benalcázar. CNN-LSTM and post-processing for EMG-based hand gesture recognition. *Intelligent Systems with Applications*, 22:200352, June 2024. ISSN 2667-3053. doi: 10.1016/j.iswa.2024.200352. URL `http://dx.doi.org/10.1016/j.iswa.2024.200352`.

[20] Gwo-Ching Chang, Wen-Juh Kang, Jer-Junn Luh, Cheng-Kung Cheng, Jin-Shin Lai, Jia-Jin J. Chen, and Te-Son Kuo. Real-time implementation of electromyogram pattern recognition as a control command of man-machine interface. *Medical Engineering and Physics*, 18(7): 529–537, 1996. ISSN 1350-4533. doi: https://doi.org/10.1016/1350-4533(96)00006-9. URL `https://www.sciencedirect.com/science/article/pii/1350453396000069`.

[21] Zeeshan O Khokhar, Zhen G Xiao, and Carlo Menon. Surface EMG pattern recognition for real-time control of a wrist exoskeleton. *BioMedical Engineering OnLine*, 9(1), August 2010. ISSN 1475-925X. doi: 10.1186/1475-925x-9-41. URL `http://dx.doi.org/10.1186/1475-925X-9-41`.

[22] Susanna Yu. Gordleeva, Sergey A. Lobov, Nikita A. Grigorev, Andrey O. Savosenkov, Maxim O. Shamshin, Maxim V. Lukoyanov, Maxim A. Khoruzhko, and Victor B. Kazantsev. Real-time EEG–EMG human–machine interface-based control system for a lower-limb exoskeleton. *IEEE Access*, 8:84070–84081, 2020. ISSN 2169-3536. doi: 10.1109/access.2020.2991812. URL `http://dx.doi.org/10.1109/ACCESS.2020.2991812`.

[23] Dezhen Xiong, Daohui Zhang, Yaqi Chu, Yiwen Zhao, and Xingang Zhao. Intuitive human-robot-environment interaction with emg signals: A review. *IEEE/CAA Journal of Automatica Sinica*, 11(5):1075–1091, May 2024. ISSN 2329-9274. doi: 10.1109/jas.2024.124329. URL `http://dx.doi.org/10.1109/JAS.2024.124329`.

[24] Levi J. Hargrove, Ann M. Simon, Aaron J. Young, Robert D. Lipschutz, Suzanne B. Finucane, Douglas G. Smith, and Todd A. Kuiken. Robotic leg control with EMG decoding in an amputee with nerve transfers. *New England Journal of Medicine*, 369(13):1237–1242, September 2013. ISSN 1533-4406. doi: 10.1056/nejmoa1300126. URL `http://dx.doi.org/10.1056/NEJMoa1300126`.

[25] Andrea Cimolato, Josephus J. M. Driessen, Leonardo S. Mattos, Elena De Momi, Matteo Laffranchi, and Lorenzo De Michieli. EMG-driven control in lower limb prostheses: a topic-based systematic review. *Journal of NeuroEngineering and Rehabilitation*, 19(1), May 2022.

ISSN 1743-0003. doi: 10.1186/s12984-022-01019-1. URL http://dx.doi.org/10.1186/s12984-022-01019-1.

[26] Bahareh Ahkami, Kirstin Ahmed, Alexander Thesleff, Levi Hargrove, and Max Ortiz-Catalan. Electromyography-based control of lower limb prostheses: A systematic review. *IEEE Transactions on Medical Robotics and Bionics*, 5(3):547–562, August 2023. ISSN 2576-3202. doi: 10.1109/tmrb.2023.3282325. URL http://dx.doi.org/10.1109/TMRB.2023.3282325.

[27] Daniele Leonardis, Carmelo Chisari, Massimo Bergamasco, Antonio Frisoli, Michele Barsotti, Claudio Loconsole, Massimiliano Solazzi, Marco Troncossi, Claudio Mazzotti, Vincenzo Parenti Castelli, Caterina Procopio, and Giuseppe Lamola. An EMG-controlled robotic hand exoskeleton for bilateral rehabilitation. *IEEE Transactions on Haptics*, 8(2): 140–151, April 2015. ISSN 2334-0134. doi: 10.1109/toh.2015.2417570. URL http://dx.doi.org/10.1109/TOH.2015.2417570.

[28] Chris Wilson Antuvan and Lorenzo Masia. An LDA-based approach for real-time simultaneous classification of movements using surface electromyography. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(3):552–561, March 2019. ISSN 1558-0210. doi: 10.1109/tnsre.2018.2873839. URL http://dx.doi.org/10.1109/TNSRE.2018.2873839.

[29] Beau Crawford, Kai Miller, Pradeep Shenoy, and Rajesh Rao. Real-time classification of electromyographic signals for robotic control. In *AAAI*, volume 5, pages 523–528, 2005.

[30] C. Tepe and M.C. Demir. Real-time classification of EMG myo armband data using support vector machine. *IRBM*, 43(4):300–308, August 2022. ISSN 1959-0318. doi: 10.1016/j.irbm.2022.06.001. URL http://dx.doi.org/10.1016/j.irbm.2022.06.001.

[31] Andrés Jaramillo-Yánez, Marco E. Benalcázar, and Elisa Mena-Maldonado. Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review. *Sensors*, 20(9):2467, April 2020. ISSN 1424-8220. doi: 10.3390/s20092467. URL http://dx.doi.org/10.3390/s20092467.

[32] Chuheng Wu, S. Farokh Atashzar, Mohammad M. Ghassemi, and Tuka Alhanai. An LSTM feature imitation network for hand movement recognition from sEMG signals. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, page 1–5. IEEE, April 2025. doi: 10.1109/icassp49660.2025.10890441. URL http://dx.doi.org/10.1109/ICASSP49660.2025.10890441.

[33] Viswanath Sivakumar, Jeffrey Seely, Alan Du, Sean R Bittner, Adam Berenzweig, Anuoluwapo Bolarinwa, Alexandre Gramfort, and Michael I Mandel. emg2qwerty: A large dataset with baselines for touch typing using surface electromyography. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 91373–91389. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/a64d53074d011e49af1dfc72c332fe4b-Paper-Datasets_and_Benchmarks_Track.pdf.

[34] Sasha Salter, Richard Warren, Collin Schlager, Adrian Spurr, Shangchen Han, Rohin Bhasin, Yujun Cai, Peter Walkington, Anuoluwapo Bolarinwa, Robert Wang, et al. emg2pose: A large and diverse benchmark for surface electromyographic hand pose estimation. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

[35] Jehan Yang, Maxwell Soh, Vivianna Lieu, Douglas J Weber, and Zackory Erickson. Emgbench: Benchmarking out-of-distribution generalization and adaptation for electromyography, 2024. URL https://arxiv.org/abs/2410.23625.

[36] Mansooreh Montazerin, Elahe Rahimian, Farnoosh Naderkhani, S. Farokh Atashzar, Svetlana Yanushkevich, and Arash Mohammadi. Transformer-based hand gesture recognition from instantaneous to fused neural decomposition of high-density EMG signals. *Scientific Reports*, 13(1), July 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-36490-w. URL http://dx.doi.org/10.1038/s41598-023-36490-w.

[37] Soheil Zabihi, Elahe Rahimian, Amir Asif, and Arash Mohammadi. TraHGR: Transformer for hand gesture recognition via electromyography. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:4211–4224, 2023. ISSN 1558-0210. doi: 10.1109/tnsre.2023.3324252. URL `http://dx.doi.org/10.1109/TNSRE.2023.3324252`.

[38] Patricio J. Cruz, Juan Pablo Vásconez, Ricardo Romero, Alex Chico, Marco E. Benalcázar, Robin Álvarez, Lorena Isabel Barona López, and Ángel Leonardo Valdivieso Caraguay. A Deep Q-Network based hand gesture recognition system for control of robotic platforms. *Scientific Reports*, 13(1), May 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-34540-x. URL `http://dx.doi.org/10.1038/s41598-023-34540-x`.

[39] Di Wu, Jie Yang, and Mohamad Sawan. Transfer learning on electromyography (EMG) tasks: Approaches and beyond. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:3015–3034, 2023. ISSN 1558-0210. doi: 10.1109/tnsre.2023.3295453. URL `http://dx.doi.org/10.1109/TNSRE.2023.3295453`.

[40] Haojie Shi, Xinyu Jiang, Chenyun Dai, and Wei Chen. EMG-based multi-user hand gesture classification via unsupervised transfer learning using unknown calibration gestures. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:1119–1131, 2024. ISSN 1558-0210. doi: 10.1109/tnsre.2024.3372002. URL `http://dx.doi.org/10.1109/TNSRE.2024.3372002`.

[41] Xinhui Li, Xu Zhang, Xiang Chen, Xun Chen, and Aiping Liu. Cross-user gesture recognition from sEMG signals using an optimal transport assisted student-teacher framework. *Computers in Biology and Medicine*, 165:107327, October 2023. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2023.107327. URL `http://dx.doi.org/10.1016/j.compbiomed.2023.107327`.

[42] Ethan Eddy, Evan Campbell, Scott Bateman, and Erik Scheme. Big data in myoelectric control: large multi-user models enable robust zero-shot EMG-based discrete gesture recognition. *Frontiers in Bioengineering and Biotechnology*, 12, September 2024. ISSN 2296-4185. doi: 10.3389/fbioe.2024.1463377. URL `http://dx.doi.org/10.3389/fbioe.2024.1463377`.

[43] Bo Xue, Le Wu, Aiping Liu, Xu Zhang, Xiang Chen, and Xun Chen. Reduce the user burden of multiuser myoelectric interface via few-shot domain adaptation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:972–980, 2023. ISSN 1558-0210. doi: 10.1109/tnsre.2023.3237181. URL `http://dx.doi.org/10.1109/TNSRE.2023.3237181`.

[44] Jingxi Xu, Runsheng Wang, Siqi Shang, Ava Chen, Lauren Winterbottom, To-Liang Hsu, Wenxi Chen, Khondoker Ahmed, Pedro Leandro La Rotta, Xinyue Zhu, Dawn M. Nilsen, Joel Stein, and Matei Ciocarlie. ChatEMG: synthetic data generation to control a robotic hand orthosis for stroke. *IEEE Robotics and Automation Letters*, 10(2):907–914, February 2025. ISSN 2377-3774. doi: 10.1109/lra.2024.3511372. URL `http://dx.doi.org/10.1109/LRA.2024.3511372`.

[45] Pedro Leandro La Rotta, Jingxi Xu, Ava Chen, Lauren Winterbottom, Wenxi Chen, Dawn Nilsen, Joel Stein, and Matei Ciocarlie. Meta-learning for fast adaptation in intent inferral on a robotic hand orthosis for stroke. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, page 4693–4700. IEEE, October 2024. doi: 10.1109/iros58592.2024.10801596. URL `http://dx.doi.org/10.1109/IROS58592.2024.10801596`.

[46] Felix Schiel, Annette Hagengruber, J Vogel, and Rudolph Triebel. Incremental learning of EMG-based control commands using gaussian processes. *CoRL*, 155:1137–1146, November 2020. URL `https://proceedings.mlr.press/v155/schiel21a`.

[47] Zineng Tang, Ziyi Yang, Chenguang Zhu, Michael Zeng, and Mohit Bansal. Any-to-Any generation via composable diffusion. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL `https://openreview.net/forum?id=2EDqbSCnmF`.

[48] Jun Zhan, Junqi Dai, Jiasheng Ye, Yunhua Zhou, Dong Zhang, Zhigeng Liu, Xin Zhang, Ruibin Yuan, Ge Zhang, Linyang Li, Hang Yan, Jie Fu, Tao Gui, Tianxiang Sun, Yu-Gang Jiang, and Xipeng Qiu. AnyGPT: Unified multimodal LLM with discrete sequence modeling. In *Proceedings of the 62nd Annual Meeting of the Association for Computational*

Linguistics (Volume 1: Long Papers), pages 9637–9662, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.521. URL `https://aclanthology.org/2024.acl-long.521/`.

[49] Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. NExT-GPT: any-to-any multimodal LLM. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

[50] Liu Zhuang, Lin Wayne, Shi Ya, and Zhao Jun. A robustly optimized BERT pre-training approach with post-training. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1218–1227, Huhhot, China, August 2021. Chinese Information Processing Society of China. URL `https://aclanthology.org/2021.ccl-1.108/`.

[51] Marco E. Benalcazar, Lorena Barona, Leonardo Valdivieso, Xavier Aguas, and Jonathan Zea. EMG-EPN-612 dataset, 2020. URL `https://zenodo.org/record/4421500`.

[52] Elisa Donati. EMG from forearm datasets for hand gestures recognition, 2019. URL `https://zenodo.org/record/3194792`.

[53] Suguru Kanoga, Takayuki Hoshino, and Hideki Asoh. Semi-supervised style transfer mapping-based framework for sEMG-based pattern recognition with 1- or 2-DoF forearm motions. *Biomedical Signal Processing and Control*, 68:102817, July 2021. ISSN 1746-8094. doi: 10.1016/j.bspc.2021.102817. URL `http://dx.doi.org/10.1016/j.bspc.2021.102817`.

[54] Praahas Amin. Raw surface electromyography dataset from myo arm band, 2021. URL `https://data.mendeley.com/datasets/d4y7fm3g79/1`.

[55] Kyung Hyun Lee, Ji Young Min, and Sangwon Byun. Electromyogram-based classification of hand and finger gestures using artificial neural networks. *Sensors*, 22(1):225, December 2021. ISSN 1424-8220. doi: 10.3390/s22010225. URL `http://dx.doi.org/10.3390/s22010225`.