# Offline Evaluation Matters: Investigation of the Influence of Offline Performance of EMG-Based Neural-Machine Interfaces on User Adaptation, Cognitive Load, and Physical Efforts in a Real-Time Application

Robert M. Hinson, Joseph Berman, I-Chieh Lee, William G. Filer, and He Huang, *Fellow, IEEE*

***Abstract*— There has been controversy about the value of offline evaluation of EMG-based neural-machine interfaces (NMIs) for their real-time application. Often, conclusions have been drawn after studying the correlation of the offline EMG decoding accuracy/error with the NMI user's real-time task performance without further considering other important human performance metrics such as adaptation rate, cognitive load, and physical effort. To fill this gap, this study aimed to investigate the relationship between the offline decoding accuracy of EMG-based NMIs and user adaptation, cognitive load, and physical effort in real-time NMI use. Twelve non-disabled subjects participated in this study. For each subject, we established three EMG decoders that yielded different offline accuracy (low, moderate, and high) in predicting continuous hand and wrist motions. The subject then used each EMG decoder to perform a virtual hand posture matching task in real time with and without a secondary task as the evaluation trials. Results showed that the high-level offline performance decoders yield the fastest adaptation rate and highest posture matching completion rate with the least muscle effort in users during online testing. A secondary task increased the cognitive load and reduced real-time virtual task competition rate for all the decoders; however, the decoder with high offline accuracy still produced the highest task completion rate. These results imply that the offline performance of EMG-based NMIs provide important insight to users' abilities to utilize them and should play an important role in research and development of novel NMI algorithms.**

***Index Terms*— EMG decoding, neural machine interface, cognitive load, adaptation.**

Manuscript received 20 February 2023; revised 11 June 2023 and 3 July 2023; accepted 12 July 2023. Date of publication 20 July 2023; date of current version 28 July 2023. This work was supported by the National Institutes of Health (NSF) under Grant 1856441 and Grant 2221479. *(Corresponding author: I-Chieh Lee.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of North Carolina at Chapel Hill Institutional Review Board under Application No. 16-0798, in March 2022.

Robert M. Hinson, I-Chieh Lee, and He Huang are with the Joint Department of Biomedical Engineering, University of North Carolina, Chapel Hill, NC 27599 USA, and also with the Joint Department of Biomedical Engineering, North Carolina State University, Raleigh, NC 27695 USA (e-mail: ichieh322@gmail.com).

Joseph Berman is with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695 USA.

William G. Filer is with the Department of Physical Medicine and Rehabilitation, UNC Chapel Hill, Chapel Hill, NC 27599 USA.

Digital Object Identifier 10.1109/TNSRE.2023.3297448

## I. Introduction

ELECTROMYOGRAPHIC (EMG) signals represent the efferent neural control commands of body movement and, therefore, have been commonly used as a neural control source for design of neural-machine interfaces (NMIs) [1]. These NMIs usually consist of an EMG sensor interface, an EMG decoding algorithm, and external virtual or physical devices. Neural signals from targeted muscles are extracted via EMG sensors placed on the skin surfaces [2] or implantable intramuscular electrodes [3]. An EMG decoding algorithm then interprets the EMG signals, determining user intent such as discrete motion classes (e.g., open/close hand) [4] or continuous joint motion [5]. To identify discrete motion classes, pattern recognition has been the most used computational framework for EMG decoder design [6]. This framework involves the extraction of features [7] from acquired EMG signals and a classification algorithm (ranging from simple linear discriminant analysis [8] to deep learning [9]) to identify the motion class. For continuous decoders, data-driven regression methods, such as artificial neural networks (ANNs) [10], [11], [12], [13], [14], [15], [16], musculoskeletal models [17], [18], [19], [20], [21], [22], [23], [24], or a combination of machine learning and a musculoskeletal model [24] have been reported to successfully predict continuous joint kinematics or kinetics. Finally, the outputs of the decoder are fed to an external machine, such as virtual reality [25], [26] or prosthetic limbs [21], [24], [27] to enable intuitive control of these machines. For example, continuous kinematics decoders have been used for position control of virtual and prosthetic joints to produce natural, multi-joint coordinated arm motions [17], [18], [20], [24].

The EMG-based NMI design procedure is often composed of offline design and analysis and real-time application and evaluation [28]. Offline design and analysis aim to ensure the best performance of the EMG decoding algorithm before its human-in-the-loop real-time applications. In the offline design phase, EMG signals and joint motion classes or kinematics are first simultaneously collected. The known input (EMG signals) and output (motion class or kinematics) data are used to train the parameters in the EMG decoders to achieve the desired offline performance. The offline performance is usually quantified by confusion matrices and classification accuracy in the case of pattern recognition for recognizing the discrete motion class [9], or correlation, root mean square error (RMSE), and coefficient of determination ($R^2$) for EMG decoders that predict continuous kinematics [28]. Iterative decoding algorithm design and offline evaluation are performed to optimize the offline decoder performance before real-time implementation of the decoder for control of a physical device or virtual environment.

However, the importance of offline evaluation/optimization of EMG-based NMI has recently been controversial in the community. This is because several recent studies have concluded decoder offline performance does not correlate with real-time task performance (e.g., task completion percentage, task completion time, etc.) of the users when utilizing the NMI [11], [29], [30]. Although these studies offer engrossing observations, the conclusion of these studies have been misinterpreted by the research community that improving or optimizing offline performance of EMG-based NMIs is less important [11]. As a result, when a new EMG deciphering algorithm is proposed in the field, the engineers are urged to conduct real-time evaluation of the new algorithm rather than careful offline analysis and optimization to show the value of the new method. Nevertheless, this concept is counterintuitive to engineers, as offline analysis is often used in the engineering development process to validate the performance and reliability of the algorithm itself and is cost-effective for troubleshooting before real-time applications. In addition, our recent study showed that the offline decoding accuracy in NMI has a significant effect on real-time virtual task performance in humans [31], contrary to the evidence observed previously.

The controversy around the importance of offline NMI analysis and optimization was drawn around the limited evaluation metrics. In fact, online NMI performance was solely quantified as the NMI user's task performance. Nevertheless, other metrics, such as a user's adaptation rate [32], cognitive load [33], and energetic exertion [30], are equally important, which are related to utility of EMG-based NMIs. These online performance metrics, together with the NMI user's task performance, have been studied in a limited manner across different EMG decoding mechanisms (such as pattern recognition vs. direct myoelectric control) [32] and feedback conditions [34] for upper limb prostheses. Results showed that low adaptation rate and high cognitive load in NMI users for prosthesis operation can be significant factors in the abandonment of upper limb prosthetic devices [35], [36], [37]. In addition, high energetic exertion in NMI use can lead to muscle fatigue, known to alter EMG signals [38], potentially further affecting the usability of the NMI for real-time control. Nevertheless, how the offline NMI decoding accuracy associated with these NMI user's physical and cognitive performance in real-time application has not been systematically investigated. Hence, it is essential to understand the influence of NMI accuracy evaluated offline on the user's adaptation, adaptation rate, and mental and physical efforts in real-time NMI use before making the conclusion on whether offline NMI evaluation and optimization is important.

Therefore, in this study we sought to investigate the question of whether the offline performance of EMG-based decoding algorithms influences users' capacity to adapt to the controller in a real-time task, the rate at which users can adapt to the controller, the effort users must exert during use, and the cognitive load associated with real-time use. We manipulated the training process of artificial neural network (ANN) EMG decoders to control the coefficient of determination ($R^2$ value) associated with each decoder in offline evaluation. Participants then used the trained decoders in a virtual hand posture matching task, training to steady state performance before final evaluation. Additionally, cognitive load was assessed using a dual task paradigm in which participants were instructed to perform the real-time task while also performing a secondary task [39], [40]. We hypothesized that improved offline performance of decoders would result in greater capacity to adapt, faster rates of adaptation, decreased effort, and decreased cognitive burden. This study provided new evidence, valuable to the scientific community, in understanding the importance of offline evaluation/optimization of EMG-based NMIs, as well as providing a rigorous reexamination of the relationship between offline performance and real-time task performance.

## II. Methods

### A. Subjects

The University of North Carolina at Chapel Hill Institutional Review Board reviewed and approved the experimental protocol (Protocol #16-0798; renewal approval on March 11, 2022). Informed consent was obtained from 12 subjects (6 male, 6 female, ages 22-31, right hand dominant) to participate. None of the participants reported neuromuscular disorders or any cognitive impairment.

### B. Data Acquisition and Real-Time Processing

Four bipolar EMG electrodes (Sensor SX230, Amplifier K800, Biometrics, Ladysmith, VA, USA) were placed over the *flexor digitorum superficialis* (FDS), *flexor carpi radialis* (FCR), *extensor digitorum communis* (EDC), and *extensor carpi radialis longus*(ECRL) muscles of the subject's dominant arm. Each muscle was identified via palpation and the skin over each muscle was prepared using an alcohol pad. Following placement of the EMG electrodes, 9 motion capture markers were placed over anatomic landmarks of the forearm and hand to allow for calculation of wrist and metacarpophalangeal (MCP) flexion/extension angles. Detailed marker setup can be found in our previous report [19]. Trajectories of markers were recorded using an optical motion capture system (Vicon Motion Systems Ltd., Oxford, UK). Marker and EMG data were synchronously collected at 100 Hz and 1000 Hz, respectively.

Next, subjects sat in front of a computer screen and rested their dominant arm on a table in front of them with their elbow held at approximately 45° of flexion during testing. The virtual posture matching task was displayed on the computer screen. A maximum voluntary contraction (MVC) was performed by the 4 muscles of interest and EMG data were recorded. Five 60 second trials of various motions were performed for training the artificial neural networks (ANNs) used to control the virtual hand in the posture matching task [19]. The 5 trials involved random, isolated MCP motion; random, isolated wrist motion; patterned, isolated MCP motion; patterned, isolated wrist motion; and random, simultaneous motion of both joints. The patterned motion of each joint consisted of cycling from full flexion to relaxed posture to full extension, and back to relaxed posture with the assistance of a metronome set to 1 beat per second. Movements between postures occurred on each beat allowing subjects to cycle through the pattern at a rate of 0.25 Hz. During trials involving random motion, subjects were directed to cover each joint's full range of motion while randomly moving at self-selected speeds.

Coordinate systems for the forearm, palm, and finger segments were defined using marker data according to previous work and wrist and MCP flexion/extension angles were determined via inverse kinematics. The EMG signals were enveloped by calculating the mean absolute value of a sliding 260 ms window to envelope the signals. The enveloped EMG signals from each muscle were normalized by the maximum enveloped EMG value from the corresponding muscle's MVC trial. The normalized, enveloped EMG data were down sampled to 100 Hz.

### C. ANN Training

The artificial neural network (ANN) algorithms was used to train EMG [31]. Two separate ANNs were implemented during each experimental session: one to predict wrist joint angles (flexion/extension) from EMG, and another to predict MCP flexion/extension angles from EMG. The predicted joints angles of each trained ANN and the measured joints angles was used to calculate the coefficient of determination $R^2$ as the stopping rule for training the three decoders with low, moderate and high offline performance. ANN decoders were used because we can easily control the offline decoding accuracy during the ANN training procedure. The ANNs used were non-linear autoregressive neural networks with external inputs (NARX) networks (Deep Learning Toolbox, MATLAB 2019a; MathWorks, Natick, MA). NARX networks (NNs) are recurrent neural networks commonly used in time series predictions [41], [42], which can also be used for the prediction of joint kinematics from EMG data [24]. The NNs consisted of a single hidden layer of 7 neurons and an output layer. The inputs consisted of the previous time step's predicted joint angle and the normalized, enveloped EMG data from the current timestep. Training initially occurred in an open-loop configuration, with measured joint angles as inputs, before being trained in a closed-loop configuration, receiving estimated joint angles as inputs.

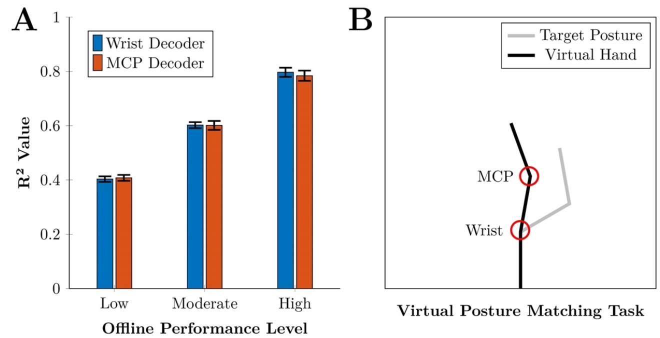


Fig. 1. (A) Summary of $R^2$ values for wrist (blue) and MCP (orange) EMG-based decoders for all subjects for low, moderate, and high offline performance levels. (B) The virtual posture matching hand interface. The virtual hand (black) was controlled by subjects to match target postures, like the one shown in gray.

Performance of each NN was evaluated by calculating the coefficient of determination, as defined in Eqn. 1:

$$R^2 = 1 - \frac{\sum_i \left(\hat{\theta}_i - \theta_i\right)^2}{\sum_i \left(\theta_i - \bar{\theta}\right)^2} \tag{1}$$

where $\bar{\theta}$, $\theta_i$, and $\hat{\theta}_i$ are the average measured joint angle, the measured joint angle at timestep i, and the estimated joint angle at timestep i, respectively. To test a large range of offline performance levels, NNs were trained to performance levels of were $R^2 \approx 0.4$, (low offline performance), $R^2 \approx 0.6$ (moderate offline performance), and $R^2 \approx 0.8$ (high offline performance). In each session, the NN predicting wrist joint angles and the NN predicting MCP joint angles were trained to the same level of offline accuracy.

The stopping criterion of the NN training was mean-square error (MSE), so the $R^2$ value of each NN was set by solving Eqn. 1 for the MSE corresponding to the desired $R^2$ value as shown in Eqn. 2:

$$MSE = \frac{\sum_i \left(\hat{\theta}_i - \theta_i\right)^2}{N} = \frac{\sum_i \left(\theta_i - \bar{\theta}\right)^2}{N}(1 - R^2) \tag{2}$$

where N represents the number of training data timesteps, and all other variables are as defined above. Arbitrary 20 second windows of data from each of the 5 trials described above were used in training each NN. A 5-fold cross-validation was done to determine performance, with 20% of each training trial withheld for testing and the other 80% used for training. Training was re-initialized and repeated until performance was $\pm 0.05$ of the target $R^2$ value. The offline performance of all NNs used by all subjects in the experiment is summarized in Fig. 1.

### D. Virtual Posture Matching Task

A 2-DoF, planar stick figure hand was displayed and visually updated at 20Hz on a computer screen for subjects to control in the virtual posture matching task [20]. Subjects were given 20 seconds to successfully match each displayed target posture by moving the wrist and MCP joints to be within ±5°

(3.6% and 5.6% of wrist and MCP task space, respectively) of their target posture values for 0.5 consecutive seconds. The target posture turned from gray to green to indicate the subject was within the target tolerance.

Testing for each subject was completed over 4 sessions. In the first session, subjects performed the virtual posture matching task using the joint angles calculated in real-time via inverse kinematics, allowing the virtual hand to perfectly mirror the subject's desired motion (an optimal performance). We estimated that the delay due to the sliding window could be estimated as approximately 50% of the length of the window, while all other processing delays were considered negligible. Since the 260 ms sliding EMG windows for the ANN decoders was estimated to cause a 130 ms delay to display the joint angles on the screen, a 130 ms delay was added to the inverse kinematic (IK) controller to simulate the delay associated with EMG-based control. Subjects were instructed to match as many targets as was possible and to do so as quickly and as accurately as possible. First, 10 practice trials were performed with 9 target postures per trial to measure adaptation. Upon completion of the practice trials, a line was fit to the final 3 trials' target completion percentages and target completion times to determine if the subject had reached steady-state performance. If the slope of this line was determined to be statistically different from 0 using a student's t-test, additional practice trials were completed until performance ceased improving. Next, 5 evaluation trials were performed with 36 target postures per trial. Finally, an additional 5 evaluation trials with 36 target postures were performed. During the evaluation trials, a secondary task was implemented to increase the cognitive load of the participants. Half of the subjects simultaneously performed the secondary task in the first set of evaluation trials and the other half performed the secondary task in the second set of evaluation trials to counteract the order effect of testing. Practice trials consisted of 9 target postures to prevent fatigue before evaluation trials with 36 target postures. Marker and EMG data were recorded for all attempted postures of each trial. The virtual hand interface and an example target posture are shown in Fig. 1B.

Following completion of the IK controller session, each subsequent session was performed using trained ANNs from each of the 3 approximate $R^2$ levels. The order in which subjects utilized each of the 3 decoder performance levels were counterbalanced so each possible permutation of decoder order was implemented for 2 subjects.

### E. Secondary Task

During 5 of the evaluation trials subjects were instructed to perform a secondary task simultaneously with the virtual posture matching task. The secondary task involved subjects attempting to spell English words backwards. Difficulty of spelling ranged from education grade 5 to grade 10 levels. Subjects were instructed to spell backward with as much accuracy and as many words as possible while performing the virtual postural matching task.

### F. Task Performance Metrics

Four performance metrics were used to evaluate subjects' performance during the virtual posture matching task: completion percentage, normalized completion time, path efficiency, and number of target posture overshoots [11], [28]. (1) Compltion percentage (%) is defined as the percent of all possible target postures matched by the subject. (2) Normalized completion time (s/rad) is defined as the duration required to match a target posture normalzied by the shortest path distance between the starting and target postures. (3) Number of overshoots is the number of times the virtual hand enters and leaves the target posture tolerance window before matching the target. (4) Path efficiency (%) is defined as the ratio of the shortest path length between the starting and target postures to the path length the virtual hand traveled to match the target posture.

### G. Adaptation Analysis

The effects of practice with controllers of all levels of offline performance were quantified in two ways: (1) the total change in performance and (2) the initial adaptation rate. The total change in performance was determined by calculating the performance metrics for the first and final practice trials and calculating the difference between the two trials for all subjects. This was calculated for all levels of offline decoder performance.

Initial adaptation rates for all subjects were determined for all metrics and were calculated for all levels of controller offline performance. An exponential curve was fit to the practice trial data of each subject using each decoder using Equation 3:

$$f(x) = Ae^{Bx} + C \tag{3}$$

where $x$ is the practice trial number, $f(x)$ is the predicted performance metric value, and A, B, and C are constants determined by the curve fitting [43]. Representative examples of practice trial data and the fit curves for completion percent are shown in Fig. 3. The derivative of the fit curve was then calculated and evaluated at the first trial ($x = 1$) to determine the initial adaptation rate.

### H. Measurement of Intuitiveness

In this study, we quantified the intuitiveness of NMI in real-time operation as the normalized error, defined as the absolute difference between subjects' physical hand joint angles and virtual hand joint angles for successfully matched target postures, normalized by each joint's range of motion, as shown below.

$$Normalized\ Error = \frac{|\theta_{virtual} - \theta_{physical}|}{RoM} \tag{4}$$

wherein, $\theta_{virtual}$ is the joint angle of the virtual hand upon successfully matching a target posture, $\theta_{physical}$ is the corresponding joint angle of the subject's physical hand upon successfully matching a target posture, and $RoM$ is the joint's range of motion [28]. The normalized error of each joint was then regressed on controller accuracy ($R^2$ value) to

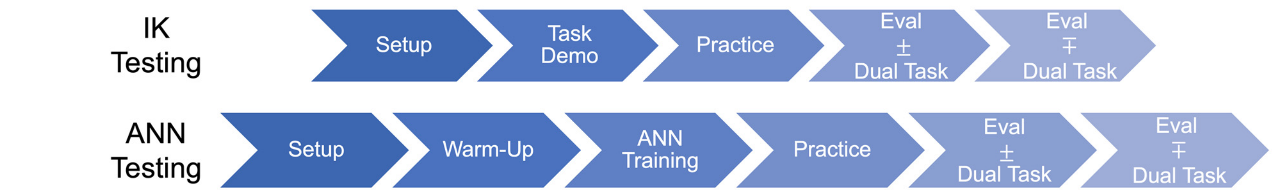


Fig. 2. A summary of the experimental workflow. First, subjects performed testing with the inverse kinematic (IK) controller (top). Subjects were prepared, had the task demonstrated, performed 10 practice trials, performed 5 evaluation trials with or without the secondary task, and then performed 5 evaluation trials without or with the secondary task. Testing with artificial neural network (ANN) controllers for all offline performance levels was the same as the IK controller testing, with the addition of ANN training data collection.

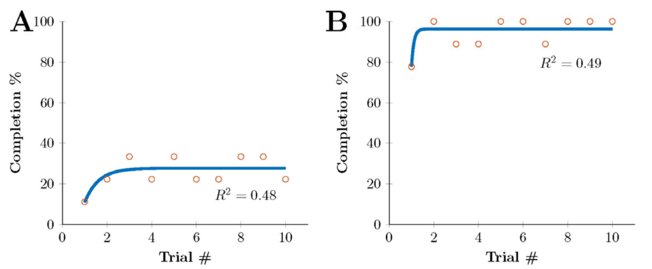


Fig. 3. Representative examples of exponential curves fit to subjects' target completion percentage data demonstrating slower (A) and faster (B) initial adaptation rates.

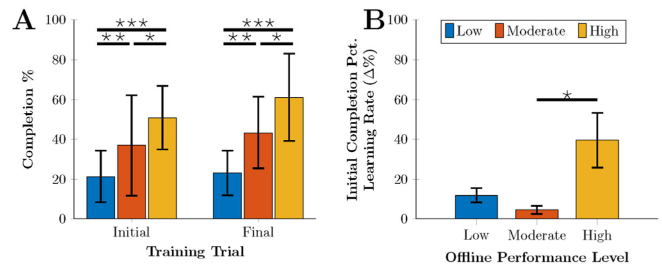


Fig. 4. (A) The change in target completion percentage from the beginning (Initial) to the end (Final) of the training/practice trials. Higher offline performance leads to higher completion percentages. However, there was no significant increase from the initial practice trial to the final training trial across all offline performance levels. (B) The initial adaptation rate for improving completion percentage was significantly higher for high offline performance decoders compared to the moderate decoders (p=0.013). Standard error bars are shown in (B). * — p<0.05; ** — p<0.01; *** — p<0.001.

determine if any relationship was present. The smaller the normalized error, the more closely matched between the user's proprioception and the NMI estimated virtual arm posture (i.e., the more intuitive the NMI control to the users).

### I. Cognitive Burden Analysis

The effect of the secondary task load on performing the virtual postural matching task was quantified by calculating the relative change in target completion percentage between the single and dual tasks. The relative change (RC) calculation is shown below.

$$RC = \frac{CP_{dual} - CP_{single}}{CP_{single}} \times 100\%, \tag{5}$$

where $CP_{dual}$ and $CP_{single}$ represent the completion percentage during the dual and single task evaluation trials, respectively [40].

### J. User Effort Analysis

Subjects' effort using each controller was quantified by calculating the normalized EMG magnitude of each muscle during the single task evaluation trials, averaged across the time required to complete the target postures [38].

### K. Statistical Analysis

A chi-square goodness of fit test was applied to all results and were determined to be sufficiently normal to apply analysis of variance (ANOVA). Mixed effects ANOVA with subjects included as random effects, and offline performance levels (low, moderate, and high), tasks (single and dual), and training state (initial and final) as fixed effects was performed and were considered significant at the $\alpha = 0.05$ level. When main effects were found to be statistically significant, Tukey's honestly significant difference test was applied to perform post-hoc analysis. All results are reported as mean ± standard deviation unless otherwise stated.

## III. Results

### A. Adaptation and Adaptation Rate

Increases in offline performance, on average, resulted in larger increases in task completion percentage over the course of the practice trials. Two of the 12 subjects required 13 practice trials to achieve steady-state performance using the high and low offline performance levels, respectively, while one subject required 11 practice trials using the moderate offline performance level. In 10 practice trials, all other subject-controller pairs achieved steady-state performance. The completion percentages at the beginning of training and following completion of training are shown in Fig. 4A across three different offline performance levels. Offline performance level was found to significantly influence performance during training ($p<0.001$) with pairwise results summarized in Fig. 4A. However, no significant difference was observed between initial performance during training, and final performance at the end of training ($p=0.104$). Initial ANOVA results showed no significant interaction ($p=0.667$) between offline performance level and training state (initial or final) and was excluded from the final model.

While there was no significant difference in total change in completion percentage, offline performance had a significant

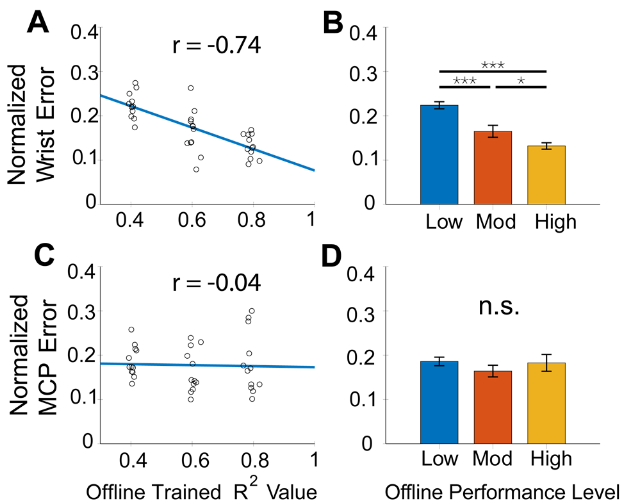


Fig. 5. The average difference between physical hand and virtual hand joint angles for the wrist (A-B) and MCP (C-D) upon successfully matching a target, normalized by joint range of motion. Higher normalized error indicates subjects' physical hands were farther away from the posture of the virtual hand. As offline performance improved, error for the wrist joint reduced while it remained unchanged for the MCP. * — p<0.05; ** — p<0.01; *** — p<0.001; n.s. – not significant.

effect on the subjects' initial adaptation rate of completion percentage (*p=0.012*) (see Fig. 4B). The high offline performance level showed significantly higher initial adaptation rates compared to the moderate level (*p=0.013*) but not the low offline performance level (*p=0.054*). No significant difference was observed between the low and moderate groups (*p=0.785*).

### B. Intuitiveness

The level of intuitiveness of each controller performance level was quantified by finding the error between the subject's physical posture and virtual hand posture upon successfully matching a target posture. The normalized wrist joint error showed strong negative correlation (r $= -0.74$) with offline performance ($R^2$ value), indicating the virtual and physical wrist joints were behaving more similarly (Fig. 5A). Normalized wrist error for the low, moderate, and high offline performance levels were shown in Fig. 5B. Tukey comparison detected significant differences between all offline performance levels (low-moderate: $p<0.001$; low-high: $p<=0.001$; moderate-high: $p=0.035$).

While controller offline performance demonstrated a strong effect on normalized wrist error when successfully matching a target posture, no correlation between $R^2$ value and normalized MCP joint error was detected (r $= -0.04$, Fig. 5C). Additionally, the groups with low, moderate, and high offline performance showed similar normalized MCP error (*p=0.420*), as shown in Fig. 5D.

### C. Effects of Cognitive Load

The effect of increased cognitive load during the dual task evaluation trials was determined using the relative change (Eqn. 5) from subjects' performance during each task individually. The dual task reduced the performance of the primary task regardless of the offline decoding performance level. The dual task reduced the performance of the primary task regardless of the offline decoding performance level. However, there is no significant difference between the three offline controllers in the secondary task (p $=$ 0.341) (Fig. 6A). In addition, increased offline performance resulted in a higher task completion percentage regardless of single or dual task conditions (Fig. 6B). Mixed effect two-way ANOVA with offline performance level and task condition (single vs. dual task) as fixed effects and subject as a random effect was performed as well and found no significant interaction between offline performance level and task condition (*p=0.997*).

### D. User Effort

Offline controller performance exhibited a significant influence on the average EMG magnitude of subjects' ECRL (*p=0.001*), FCR (*p<0.001*), and FDS (*p=0.003*) muscles, but not EDC (*p=0.220*). All muscle activation results across different decoders, including the IK method (i.e., the virtual hand was driven by actual hand motion), are summarized in Fig. 7.

Compared to ECRL activations while performing the posture matching task with the IK controller (0.07 $\pm$ 0.04), average ECRL activations were significantly higher using controllers with low (0.11 $\pm$ 0.04, *p=0.027*) and moderate (0.13 $\pm$ 0.06, *p<0.001*) offline performance. However, average ECRL activations using the controllers from the high offline performance group (0.09 $\pm$ 0.05) were not significantly higher compared to when using the IK controller (*p=0.264*).

Similar to ECRL, average activations of FCR while using the low performance (0.16 $\pm$ 0.08) and moderate performance (0.15 $\pm$ 0.06) controllers were significantly higher (*p<0.001* and *p=0.002*, respectively) compared to FCR activations while utilizing the IK controller (0.08 $\pm$ 0.03). Meanwhile, using controllers with high offline performance resulted in FCR activations (0.11 $\pm$ 0.04) not significantly different from those seen in trials using the IK controller (*p=0.220*).

Finally, average activations of FDS displayed trends similar to those observed in ECRL and FCR. During use of the IK controller, average activations of FDS were observed to be 0.07 $\pm$ 0.03. Utilizing controllers with low (0.13 $\pm$ 0.05) and moderate (0.13 $\pm$ 0.06) offline performance resulted in significantly higher activations, compared to those recorded using the IK controller (*p=0.007* and *p=0.006*, respectively). Controllers with high offline performance demonstrated average activations (0.11 $\pm$ 0.08) not significantly different from those seen using the IK controller (*p=0.120*).

## IV. Discussion

In this study, we set out to examine the relationship between EMG controller offline predictive accuracy and an individual's ability to adapt and utilize the controller online. We also explored the intuitiveness, cognitive burden, and physical effort in using EMG controllers with varying level of offline decoding accuracy. We hypothesized improved offline performance of decoders would result in greater capacity to adapt, faster rates of adaptation, decreased effort, and decreased cognitive burden.

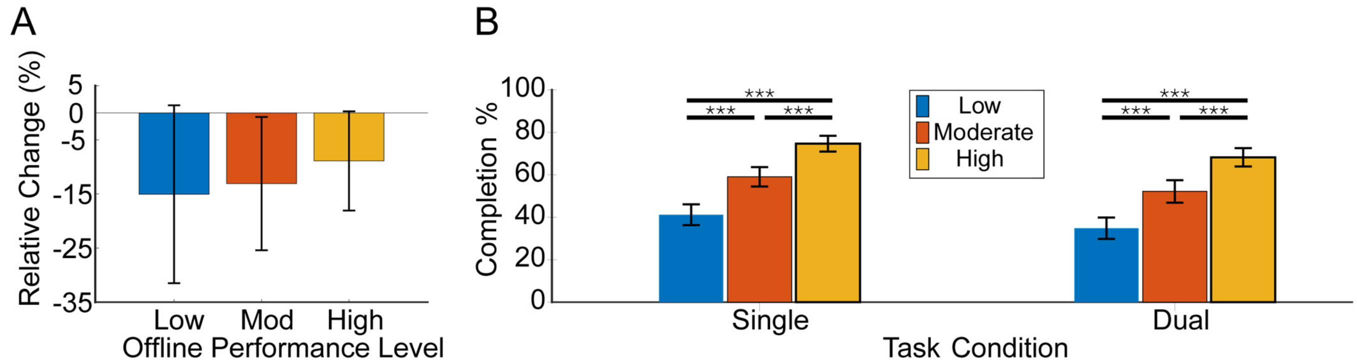


Fig. 6. (A). Effects from increased cognitive load as shown by the relative change from single task target posture completion percentage caused by the introduction of the secondary task with 95% confidence intervals. Negative relative changes indicate decreased performance. No significant difference was found between all levels. (B). Average completion percentages for the single and dual task setups are shown with standard error bars. As offline performance improves, completion percentage increases in both cases. *** —$p<0.001$.

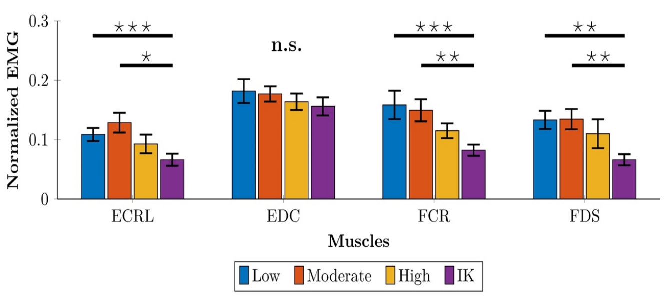


Fig. 7. Subjects' effort levels, shown by average muscle activations during the single task evaluation trials. No significant differences were seen in the average activation of EDC across all control schemes used for posture matching. However, for ECRL, FCR, and FDS, the EMG controllers with low (blue) and moderate (orange) offline performance demonstrated significantly higher muscle activations compared to the IK controller (purple). Meanwhile the EMG controllers with high (yellow) offline performance showed similar activation levels to the IK controller, indicating less effort was necessary using EMG controllers with higher offline performance. Standard error bars are displayed above. * — $p<0.05$; ** — $p<0.01$; *** — $p<0.001$; n.s. – not significant.

### *A. Users Adapt to EMG-Based NMI Faster When the Offline Decoding Accuracy Increase*

The virtual posture matching task performance metrics can be broken down into two categories: overall capability to perform the task, represented by completion percentage, and efficiency with which the task is completed successfully, represented by normalized completion time, path efficiency, and number of target overshoots.

As offline performance improved (as evidenced by higher $R^2$ values), subjects' capacity to learn and improve their target completion percentage increased (Fig. 4A). While the trend in capacity to improve completion percentage was evident, the magnitudes of these changes were not significant. However, the controllers with the highest offline performance demonstrated faster initial adaptation rates for improving completion percentage (Fig. 4B). The controllers with high offline performance demonstrated significantly higher initial adaptation rates compared to the moderate level, and approached significance compared to the low level. These results indicate once offline performance reaches a sufficiently high level, users can more quickly and intuitively figure out what they can perform in the task space with the given interface. One of the reasons is that the improved offline performance led the virtual hand to mimic the desired physiologic motion of the user more closely (i.e., better intuitiveness), as shown in Fig. 5, indicating the controller is more intuitive. This indicates that an EMG-based NMI with high offline decoding accuracy enables more intuitive control (i.e., the NMI identified user intent is more consistent with the user's internal model of motion [44]), a potential factor making the controllers easier to use without much effort in learning/adapting.

### *B. Effect of Dual Task on Task Performance Using Different Levels of Accuracy of Offline Decoding*

According to the capacity sharing model of the cognitive-motor dual task paradigm, individuals have a limited total cognitive capacity [37], [38]. The cognitive domains required increase proportionally with the number of concurrent tasks. If the requirements of the task exceed the total capacity, performance will be affected. Hence, we hypothesized that improved offline decoding accuracy would require less cognitive effort to control the virtual hand, resulting in less performance impact when a secondary task was introduced. Although the highest offline performing NMI did yield the smallest relative change (i.e., cost of secondary task), unfortunately, the result did not reach statistical significance. One potential reason is that the designed secondary task of backward spelling might be too difficult for some subjects. In future work, including additional secondary tasks in the testing may reduce the variations of individuals in performing the secondary task to assess the cognitive workload. In addition, other objective measures, such as heart rate, electroencephalography, eye tracking measures, detection response task, and NASA TLX [32], [45], during the postural matching task performance can be explored to quantify the cognitive load. Finally, regardless of the single or dual task scenario, more accurate offline decoding resulted in higher primary task performance, indicating the importance in maintaining high offline decoding accuracy for real life NMI applications.

### C. User Effort Reduces With Increased of Offline Decoding Accuracy

The magnitude of normalized EMG in IK controllers was significantly smaller than controller with low and moderate offline performance, and there was no significant difference between IK and controller with high offline performance for ECRL, FCR, and FDS muscles (three out of four muscles used to control the virtual hand) (see Fig. 7). This result indicates that as controller offline performance improved, subjects were able to better perform the virtual posture matching task (higher completion percentage) while also reducing the effort required, as the reduction of muscle activations [38]. The IK controller results offer an important reference as it depicted the ideal scenario of accomplishing the posture matching task which closely imitates the user's desired motion. The reduced effort associated with higher offline performance suggests subjects are less likely to become fatigued with prolonged use. Avoiding fatigue is important for any EMG-based controller, as fatigue results in changes to the EMG signals, which often degrade controller performance [46].

### D. Limitations

This study is not without its limitations. As previously mentioned, there is a potential need for a larger sample size to better determine the adaptation capacity and effect of increased cognitive load associated with controllers of various offline performance levels. Another limitation is the time users were required to maintain the target posture (0.5 s), while consistent with previous literature, is still a short amount of time. In the future, similar experiments may involve providing users with a fixed amount of time to achieve and maintain a target posture and calculate the percentage of time the target posture was achieved. Additionally, the observed cognitive trade-off between the two tasks seen in the dual task paradigm prevented meaningful insight into the effect of increased cognitive load on subjects while using various controllers.

## V. Conclusion

In this study, we explored the effect of EMG-based NMIs with different offline decoding accuracy levels on users' ability to adapt to the NMI controller, how users respond to increased cognitive burden, and the user's effort required to perform the posture matching task. The NMI with high offline performance level predicted hand and wrist positions closer to the physical arm posture and therefore was more intuitive to the user. The human user also showed larger capacity to improve task completion percentage with a faster initial adaptation rate through practice when using the NMI with higher offline performance for virtual hand control. Interestingly, subjects can benefit the most from additional training and practice in using the NMI with moderate offline accuracy. Improved offline performance was also associated with lower effort by users, allowing for less fatigue and better controller stability. Finally, a secondary task took additional cognitive resources away from the user, leading to a similar amount of reduction for online NMI performance regardless of the offline performance level. However, for both single and dual task tests, the NMI with high offline performance level yielded the best online task completion percentage. In summary, our results implied that offline performance of EMG-based NMIs can be indicative of overall capability of users in using NMI online for external device control. Therefore, offline analysis and iterative optimization of EMG decoding accuracy offline is essential to making EMG-based NMI design useful and acceptable by the users for different applications.

## References

[1] D. Farina et al., "The extraction of neural information from the surface EMG for the control of upper-limb prostheses: Emerging avenues and challenges," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 4, pp. 797–809, Jul. 2014.

[2] C. J. De Luca, "The use of surface electromyography in biomechanics," *J. Appl. Biomech.*, vol. 13, no. 2, pp. 135–163, May 1997.

[3] R. F. Weir, P. R. Troyk, G. A. DeMichele, D. A. Kerns, J. F. Schorsch, and H. Maas, "Implantable myoelectric sensors (IMESs) for intramuscular electromyogram recording," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 1, pp. 159–171, Jan. 2009.

[4] T. A. Kuiken et al., "Targeted muscle reinnervation for real-time myoelectric control of multifunction artificial arms," *JAMA*, vol. 301, no. 6, pp. 619–628, 2009.

[5] Q. Ding, J. Han, and X. Zhao, "Continuous estimation of human multi-joint angles from sEMG using a state-space model," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 9, pp. 1518–1528, Sep. 2017.

[6] A. Fougner, Ø. Stavdahl, P. J. Kyberd, Y. G. Losier, and P. A. Parker, "Control of upper limb prostheses: Terminology and proportional myoelectric control—A review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 5, pp. 663–677, Sep. 2012.

[7] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, Mar. 1993.

[8] H. Huang, T. A. Kuiken, and R. D. Lipschutz, "A strategy for identifying locomotion modes using surface electromyography," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 1, pp. 65–73, Jan. 2009.

[9] U. Côté-Allard et al., "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019.

[10] J. L. G. Nielsen, S. Holmgaard, N. Jiang, K. B. Englehart, D. Farina, and P. A. Parker, "Simultaneous and proportional force estimation for multifunction myoelectric prostheses using mirrored bilateral training," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 681–688, Mar. 2011.

[11] N. Jiang, I. Vujaklija, H. Rehbaum, B. Graimann, and D. Farina, "Is accurate mapping of EMG signals on kinematics needed for precise online myoelectric control?" *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 549–558, May 2014.

[12] L. Pan, D. L. Crouch, and H. Huang, "Comparing EMG-based human–machine interfaces for estimating continuous, coordinated movements," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2145–2154, Oct. 2019.

[13] S. Muceli and D. Farina, "Simultaneous and proportional estimation of hand kinematics from EMG during mirrored movements at multiple degrees-of-freedom," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 371–378, May 2012.

[14] P. Xia, J. Hu, and Y. Peng, "EMG-based estimation of limb movement using deep learning with recurrent convolutional neural networks," *Artif. Organs*, vol. 42, no. 5, pp. E67–E77, May 2018.

[15] 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.

[16] J. Liu, S. H. Kang, D. Xu, Y. Ren, S. J. Lee, and L.-Q. Zhang, "EMG-based continuous and simultaneous estimation of arm kinematics in able-bodied individuals and stroke survivors," *Frontiers Neurosci.*, vol. 11, Aug. 2017, Art. no. 480.

[17] D. Blana, E. K. Chadwick, A. J. van den Bogert, and W. M. Murray, "Real-time simulation of hand motion for prosthesis control," *Comput. Methods Biomech. Biomed. Eng.*, vol. 20, no. 5, pp. 540–549, Apr. 2017.

[18] D. Blana et al., "Model-based control of individual finger movements for prosthetic hand function," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 3, pp. 612–620, Mar. 2020.

[19] D. L. Crouch and H. Huang, "Lumped-parameter electromyogram-driven musculoskeletal hand model: A potential platform for real-time prosthesis control," *J. Biomech.*, vol. 49, no. 16, pp. 3901–3907, Dec. 2016.

[20] L. Pan, D. L. Crouch, and H. Huang, "Myoelectric control based on a generic musculoskeletal model: Toward a multi-user neural-machine interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 7, pp. 1435–1442, Jul. 2018.

[21] M. Sartori, G. Durandau, S. Došen, and D. Farina, "Robust simultaneous myoelectric control of multiple degrees of freedom in wrist-hand prostheses by real-time neuromusculoskeletal modeling," *J. Neural Eng.*, vol. 15, no. 6, Dec. 2018, Art. no. 066026.

[22] M. Sartori, D. G. Llyod, and D. Farina, "Neural data-driven musculoskeletal modeling for personalized neurorehabilitation technologies," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 5, pp. 879–893, May 2016.

[23] W. Wu, K. R. Saul, and H. Huang, "Using reinforcement learning to estimate human joint moments from electromyography or joint kinematics: An alternative solution to musculoskeletal-based biomechanics," *J. Biomech. Eng.*, vol. 143, no. 4, Apr. 2021.

[24] J. Berman, R. Hinson, I.-C. Lee, and H. Huang, "Harnessing machine learning and physiological knowledge for a novel EMG-based neural-machine interface," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 4, pp. 1125–1136, Apr. 2023.

[25] M. S. P. de Melo, J. G. da Silva Neto, J. M. X. N. Teixeira, A. E. F. Da Gama, and V. Teichrieb, "An EMG-based virtual reality application for motor rehabilitation," in *Proc. 21st Symp. Virtual Augmented Reality (SVR)*, Oct. 2019, pp. 170–177, doi: 10.1109/SVR.2019.00041.

[26] J. W. Yoo, D. R. Lee, Y. J. Sim, J. H. You, and C. J. Kim, "Effects of innovative virtual reality game and EMG biofeedback on neuromotor control in cerebral palsy," *Biomed. Mater. Eng.*, vol. 24, no. 6, pp. 3613–3618, 2014.

[27] D. Blana, J. G. Hincapie, E. K. Chadwick, and R. F. Kirsch, "A musculoskeletal model of the upper extremity for use in the development of neuroprosthetic systems," *J Biomech*, vol. 41, no. 8, pp. 1714–1721, 2008.

[28] D. Farina et al., "Toward higher-performance bionic limbs for wider clinical use," *Nature Biomed. Eng.*, vol. 7, no. 4, pp. 473–485, May 2021.

[29] L. Hargrove, Y. Losier, B. Lock, K. Englehart, and B. Hudgins, "A real-time pattern recognition based myoelectric control usability study implemented in a virtual environment," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2007, pp. 4842–4845.

[30] A. Krasoulis, S. Vijayakumar, and K. Nazarpour, "Effect of user practice on prosthetic finger control with an intuitive myoelectric decoder," *Frontiers Neurosci.*, vol. 13, Sep. 2019, Art. no. 480.

[31] R. M. Hinson, J. Berman, W. Filer, D. Kamper, X. Hu, and H. Huang, "Offline evaluation matters: Investigation of the influence of offline performance on real-time operation of electromyography-based neural-machine interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 680–689, 2023.

[32] M. M. White et al., "Usability comparison of conventional direct control versus pattern recognition control of transradial prostheses," *IEEE Trans. Hum.-Mach. Syst.*, vol. 47, no. 6, pp. 1146–1157, Dec. 2017.

[33] M. Zahabi et al., "Application of cognitive task performance modeling for assessing usability of transradial prostheses," *IEEE Trans. Hum.-Mach. Syst.*, vol. 49, no. 4, pp. 381–387, Aug. 2019.

[34] J. Gonzalez, H. Soma, M. Sekine, and W. Yu, "Psycho-physiological assessment of a prosthetic hand sensory feedback system based on an auditory display: A preliminary study," *J. NeuroEng. Rehabil.*, vol. 9, no. 1, p. 33, 2012.

[35] E. Biddiss and T. Chau, "Upper-limb prosthetics: Critical factors in device abandonment," *Amer. J. Phys. Med. Rehabil.*, vol. 86, no. 12, pp. 977–987, Dec. 2007.

[36] E. A. Biddiss and T. T. Chau, "Upper limb prosthesis use and abandonment: A survey of the last 25 years," *Prosthetics Orthotics Int.*, vol. 31, no. 3, pp. 236–257, Sep. 2007.

[37] D. S. Childress, "Closed-loop control in prosthetic systems: Historical perspective," *Ann. Biomed. Eng.*, vol. 8, nos. 4–6, pp. 293–303, Jul. 1980.

[38] A. Luttmann, M. Jäger, and W. Laurig, "Electromyographical indication of muscular fatigue in occupational field studies," *Int. J. Ind. Ergonom.*, vol. 25, no. 6, pp. 645–660, Jul. 2000.

[39] M. Tombu and P. Jolicœur, "A central capacity sharing model of dual-task performance," *J. Experim. Psychol., Hum. Perception Perform.*, vol. 29, no. 1, pp. 3–18, 2003.

[40] P. Plummer and G. Eskes, "Measuring treatment effects on dual-task performance: A framework for research and clinical practice," *Frontiers Hum. Neurosci.*, vol. 9, p. 225, Apr. 2015.

[41] Q. Liu, W. Chen, H. Hu, Q. Zhu, and Z. Xie, "An optimal NARX neural network identification model for a magnetorheological damper with force-distortion behavior," *Frontiers Mater.*, vol. 7, Feb. 2020, Art. no. 10.

[42] J. M. P. Menezes and G. A. Barreto, "Long-term time series prediction with the NARX network: An empirical evaluation," *Neurocomputing*, vol. 71, nos. 16–18, pp. 3335–3343, Oct. 2008.

[43] Y.-T. Liu, G. Mayer-kress, and K. M. Newell, "Beyond curve fitting: A dynamical systems account of exponential learning in a discrete timing task," *J. Motor Behav.*, vol. 35, no. 2, pp. 197–207, Jun. 2003.

[44] M. Kawato, "Internal models for motor control and trajectory planning," *Current Opinion Neurobiol.*, vol. 9, no. 6, pp. 718–727, Dec. 1999.

[45] M. D. Paskett et al., "Improving upper-limb prosthesis usability: Cognitive workload measures quantify task difficulty," *medRxiv*, Aug. 2022.

[46] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *J. NeuroEng. Rehabil.*, vol. 7, no. 1, pp. 1–13, Dec. 2010.