Deep Reinforcement Learning for Control of Time-Varying Musculoskeletal Systems With High Fatigability: A Feasibility Study

Jessica Abreu, Douglas C. Crowder, and Robert F. Kirsch, *Member, IEEE*

***Abstract*— Functional electrical stimulation (FES) can be used to restore motor function to people with paralysis caused by spinal cord injuries (SCIs). However, chronically-paralyzed FES-stimulated muscles can fatigue quickly, which may decrease FES controller performance. In this work, we explored the feasibility of using deep neural network (DNN) controllers trained with reinforcement learning (RL) to control FES of upper-limb muscles after SCI. We developed upper-limb biomechanical models that exhibited increased muscle fatigability, decreased muscle recovery, and decreased muscle strength, as observed in people with chronic SCIs. Simulations confirmed that controller training time and controller performance are impaired to varying degrees by muscle fatigability. Also, the simulations showed that large muscle strength asymmetries between opposing muscles can substantially impair controller performance. However, the results of this study suggest that controller performance for highly-fatigable musculoskeletal systems can be preserved by allowing for rest between movements. Overall, the results suggest that RL can be used to successfully train FES controllers, even for highly-fatigable musculoskeletal systems.**

***Index Terms*— Reinforcement learning, functional electrical stimulation, motor control, spinal cord injury, biomechanical model.**

I. Introduction

EVERY YEAR, over 130,000 people worldwide develop paralysis as a result of spinal cord injuries (SCIs) [1], and over half of these injuries affect mobility in all four limbs [2]. Most of the people with upper extremity paralysis indicate that regaining the ability to manipulate objects is their highest priority [3], [4]. Also, people with SCI would prefer to regain control of their own limbs, as opposed to using external assistive devices [5]. By eliciting contractions in paralyzed muscles, functional electrical stimulation (FES) can satisfy these design objectives.

Many studies have demonstrated that FES can restore some degree of upper-limb motor function to people with paralysis [6], [7], [8], [9]. People with SCIs have been able to complete functional tasks by controlling FES systems with few degrees of freedom and pre-programmed hand grasps [10]. Yet, the FES systems that have been demonstrated to date provide low-dimensional control, which falls short of approximating natural upper-limb function. Also, the implementation of FES systems requires continuous intervention from highly-skilled clinicians and engineers. Therefore, FES systems to restore upper-limb motor function to individuals with SCIs have not been widely translated into clinical practice.

If FES neuroprostheses are to be more widely used, they will require controllers that can address the aforementioned limitations. FES controllers should be easy to train, and they should maintain performance with minimal intervention from experts. Ideally, FES controllers should be effective in coordinating multiple actuators, providing multidimensional control that approximates natural movement. Since chronic paralysis causes muscles to become substantially weaker and more fatigable [11], [12], [13], [14], FES controllers should be effective even for highly fatigable and atrophied muscles. While many of these needs have been addressed individually by previous studies [15], [16], [17], there are no FES controller architectures, to the best of our knowledge, that can meet these requirements simultaneously.

Recently, deep reinforcement learning (DRL) has been used to train FES controllers that can meet many of these needs. By emulating natural learning, RL automatically adjusts controller parameters to maximize performance. Therefore, RL may prevent labor-intensive manual adjustments of controller parameters. RL has been used to control a multi-actuator biomechanical model [18], [19], an upper-limb FES system [20], and robotic arms performing complex motor tasks [21]. More recently, an RL technique called *Hindsight Experience Replay* (HER) [22] has been used to train FES controllers in as little as 15 minutes [23]. DRL controllers were able to effectively control a multi-input, multi-output biomechanical model of the arm in a large workspace, and trained controllers required minimal retraining for considerable changes in biomechanical properties [24]. However, previous studies did not consider musculoskeletal systems with highly fatigable and atrophied muscles, such as those observed

Manuscript received 1 March 2022; revised 1 July 2022; accepted 22 August 2022. Date of publication 5 September 2022; date of current version 19 September 2022. This work was supported by the National Institute for Biomedical Imaging and Bioengineering under Award T32EB004314. *(Corresponding author: Robert F. Kirsch.)*

Jessica Abreu and Robert F. Kirsch are with the Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH 44106 USA, and also with the Functional Electrical Stimulation Center, Louis Stokes Cleveland VAMC, Cleveland, OH 44106 USA (e-mail: rfk3@case.edu).

Douglas C. Crowder is with the Cognitive and Emerging Computing Department, Sandia National Laboratories, Albuquerque, NM 87123 USA.

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2022.3203970, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2022.3203970

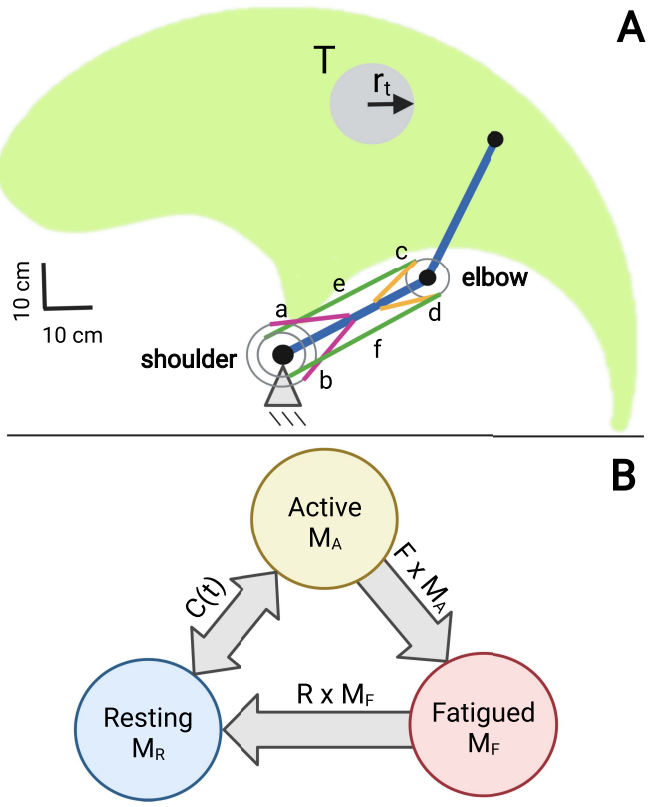

Fig. 1. **Arm Model and Motor Task. (A):** Diagram representing the musculoskeletal model of the arm and the motor task. The controller was tasked with moving the arm towards a target region *T* inside the workspace (light green). For some experiments, we chose to artificially link the fatigue levels in opposing muscles (in matching colors) to assess the effects of muscle strength asymmetries on controller performance. (a) anterior deltoid, (b) posterior deltoid, (c) brachialis, (d) short head of the triceps, (e) biceps, (f) long head of the triceps. **(B):** Visual representation of the compartment model of fatigue implemented in this work, adapted from [27]. The fatigue model was incorporated into the musculoskeletal model, allowing for the estimation of fatigue levels for each muscle during arm control.

in people with SCIs [11], [12], [13]. Fatigue is well-recognized as a major challenge for effective and robust FES control [17], [25], [26]. Because time-varying fatigue levels cannot be easily estimated in real time, they cannot be used as controller inputs. Consequently, for fatigable systems, the same control input is likely to result in time-varying muscle forces. Therefore, it is unclear if existing DRL controllers will be effective for people with SCI.

In this study, we demonstrate DRL for control of a time-varying, fatigable musculoskeletal arm model representing the biomechanical properties of people with SCIs. We explore the impact of muscle fatigue and muscle atrophy on controller performance. Also, we characterize the impact of muscle strength (i.e. maximum muscle force) asymmetries on DRL controllers. This study supports the feasibility of using DRL to train FES controllers for people with SCI.

II. Methods

A. Musculoskeletal Model

To evaluate controller performance, we used an existing musculoskeletal model of the human arm, as described previously [15], [19], [23]. Figure. 1A shows a diagram of the model. The model contained two segments representing the forearm and the upper arm. The segments were connected by 2 pin joints representing the shoulder and the elbow. The arm model included two degrees of freedom: horizontal flexion and extension of the shoulder, and flexion and extension of the elbow. The movement of the arm was constrained to a horizontal plane and the weight of the arm was supported, as if moving on a tabletop with no friction. The model included a total of 6 actuators represented by Hill muscle models [28]. Hill muscle parameters were extracted from [29], [30], and limb segment dimensions were calculated from [31] for a male subject with a height of 177 cm and weighting 80 kg, as in similar RL studies that demonstrated suitable motor behavior [19], [23] and robustness changes in biomechanical properties [24]. See Supplementary Tables I and II for an overview of key musculoskeletal parameters. As shown in Figure. 1A, 4 actuators acted on only one joint, roughly approximating the functions of the anterior deltoid (a) and the posterior deltoid (b) on the shoulder, and the functions of the brachialis (c) and the short head of the triceps (d) on the elbow. Two actuators acted on both joints, approximating the functions of the biceps (e) and the long head of the triceps (f). Simulations were performed using forward Euler approximation with model states updated every 20 ms, which has been found to provide accurate control in previous studies [15], [18], [19], [23], [24], and in preliminary simulations in the current study.

B. Modeling Fatigue

To investigate the impact of fatigability on controller performance, we implemented a previously-validated fatigue model [27], [32], [33]. The fatigue model was incorporated into the musculoskeletal model of the arm, enabling the continuous estimation of fatigue levels during arm control for each muscle. Since the added computational burden was proportional to the number of controlled muscles, we prioritized fatigue models that were reasonably accurate and computationally efficient. In spite of its simplicity, the model proposed by [27] could accurately predict fatigue levels for isometric [32] and intermittent motor tasks [33], and it did not affect simulation times in any noticeable manner in this study. Figure. 1B shows a visual representation of the fatigue model, adapted from [27]. The fatigue model included three compartments representing three possible states for motor units: (1) resting motor units M_R, (2) activated motor units M_A, and (3) fatigued motor units M_F.

Eqs. 1 describe the mathematical representation of the compartment model. $C(t)$ is a bidirectional muscle activation-deactivation drive function, as described previously [27]. R and F are the recovery and fatigue coefficients, respectively. R determines the rate at which fatigued motor units become available to perform contractions, and F determines the rate at which activated motor units become fatigued.

$$\begin{aligned}\frac{dM_R}{dt} &= -C(t) + R \times M_F \\ \frac{dM_A}{dt} &= C(t) - F \times M_A \\ \frac{dM_F}{dt} &= F \times M_A - R \times M_F \end{aligned} \tag{1}$$

TABLE I
MODEL PARAMETERS (SEE SECTION II-C)

Model	R	F	Maximum Muscle Force (%)
Avg. Male	0.0051	0.0111	100
Chronic SCI	0.0018	0.0323	30
SCI Exercised	0.0034	0.0217	50

Table I summarizes the fatigue parameters of the models implemented in this work. We used hand grip force data available in [34] to estimate R and F coefficients representing a healthy person with no SCI. The hand grip data was digitized from Figure. 9 [34], and fit to the fatigue compartment model using a least squares regression to extract R and F. To estimate the R and F coefficients for people with chronic SCIs, we used plantar flexion torque data available from a study investigating fatigue in the soleus muscle of people with SCI [12]. The torque data was digitized and fit to the fatigue compartment model to extract R and F, as described above. Simulations that use the R and F coefficients obtained from [12] should be considered conservative estimators of DRL performance, because current FES systems use FES-induced exercise to strengthen the muscles and decrease fatigability [35].

Exercise leads to increased blood flow to the muscles, which is likely to result in faster recovery from fatigue [33]. Also, previous studies demonstrated that FES-induced exercise can substantially reduce the fatigability of paralyzed muscles [35]. To model the effect of long-term exercise on the fatigability and recovery of paralyzed muscles, we averaged the R and F coefficients between the healthy model and the model representing chronic SCI, as shown in Table I (see Supplementary Figures. 4 and 5 for a sensitivity analysis on the R and F coefficients). For all conditions, fatigue was applied by scaling the output forces of the Hill-type actuators [23], [24].

C. Modeling Atrophy

To model muscle atrophy, we scaled the output forces of the Hill-type actuators of the *Chronic SCI* and *SCI Exercised* models, as shown in Table I. For each model, all muscles were weakened by the same percentage. The *Chronic SCI* model was scaled using previously reported atrophy levels (approximately 70%) for the triceps muscles of people with cervical SCI [13]. The *SCI Exercised* model was scaled using the ratio between the soleus torques produced by people with SCI after long-term FES exercise [35], and the soleus torques produced by average males [36]. Atrophy levels that were estimated using quantitative plantarflexion torque data agree with our own qualitative observations of upper-limb atrophy after FES exercise in people with tetraplegia [10], [37]. Anecdotally, our group observed that exercised upper extremity muscle strength in chronically paralyzed individuals is typically 50% of non-paralyzed individuals [10], [37]. To assess the impact of this assumption, we performed a sensitivity analysis measuring the effect of different levels of atrophy on controller performance for the *SCI Exercised* model (see Supplementary Figure. 6).

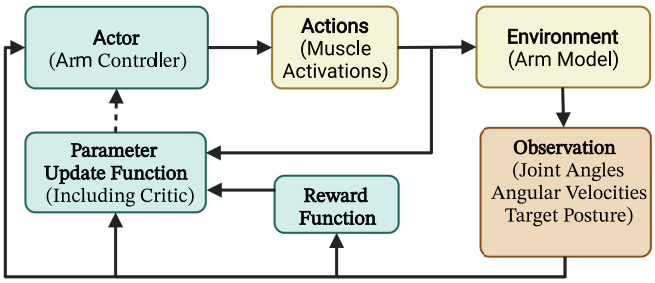

Fig. 2. **Reinforcement Learning Controller**. We used an actor-critic RL algorithm where the actor was a DNN that observed kinematic state variables and controlled muscle activations, while the critic was a DNN that mapped state-action pairs to expected rewards. The expected rewards were used by a parameter update function to adapt the actor network and maximize the rewards received during controller training.

D. DRL Controller Implementation

We used reinforcement learning (RL) [38] to train a DNN to control the muscle activations in a musculoskeletal arm model. Figure. 2 shows the implemented controller training paradigm. At each time step, the controller received the kinematic state of the system, described by the joint angular positions and joint angular velocities of the arm, as well as the target posture in angular coordinates. The action space was a 6-dimensional vector containing commanded muscle activations over a range of [0, 1] for each of the 6 actuators in the musculoskeletal model. The reinforcement learning agent was given a reward at each step according to Equation 2, where I_{at} was a boolean that was 1 if the endpoint of the arm was inside the target region T (see Figure. 1A) and 0 otherwise, and $\vec{a}$ was a 6 dimensional vector containing the muscle activations of the arm. The first term rewarded the controller for moving the arm into the target region, the second term penalized movement duration, and the third term penalized higher muscle activations to encourage lower levels of muscle fatigue. The second and the third terms were included to promote controller training convergence, as described in [23] and [24].

$$r = 1 \times I_{at} - 0.1 - 0.245 \times ||\vec{a}||_2$$
$$\vec{a} \in \mathbb{R}^6, \quad a_i \in [0, 1] \ \forall i \in \{1, 2, \ldots, 6\} \quad (2)$$

To train the arm controller, we used an actor-critic RL algorithm incorporating *Twin-Delayed Deep Deterministic Policy Gradients* (TD3) [39] and *Hindsight Experience Replay* (HER) [22], as described previously [23], [24]. Briefly, the actor observed kinematic state variables and target kinematic state variables and chose muscle activation values, as shown in Figure. 2. The critic mapped action-state pairs to expected rewards. The expected rewards were used to update the actor in order to maximize rewards. Both the actor and the critic were feedforward DNNs containing 2 layers and 64 nodes per layer.

We used the *stable-baselines3* implementation of the TD3-HER algorithm [40] (see a visual representation of

the algorithm in Supplementary Figure. 8). The RL controller was implemented using *stable-baselines3* and custom software written in Python 3.7. The musculoskeletal model was implemented in C [15]. RL hyperparameters were as described in a similar study [23], and they are available in Supplementary Table III.

E. Motor Task

For all simulations, the controller was tasked with reaching to arbitrary target locations in a 2D workspace, as shown in Figure. 1A. The targets were circular regions with radii of 7.5 cm [18], [19], [23], [24]. Target locations were randomly sampled from a continuous uniform distribution across the workspace. Since the arm was supported against gravity, the motion of the arm could be described as reaching across a frictionless tabletop or moving within a horizontal planar region using a mobile arm support [37]. The controller was given 1 second to complete the task, and most successful reaches took less than 0.4 seconds. The endpoint of the arm had to remain in the target area for 100 ms in order for the task to be considered successful. To model natural arm use, the arm started at the end state of the previous reach, as if continuously moving across the workspace.

While the theoretical joint angle ranges for the elbow and the shoulder in this model were [5°, 170°] and [−20°, 130°] respectively, these ranges included passive postures that could only be maintained with the aid of external forces. To estimate the actual workspace of the arm, we used data from previous RL controller implementations using the same arm model [23], [24]. We only included regions where previous RL controllers could acquire targets with greater than 95% probability. By choosing this workspace, we could represent controller performance for different simulation conditions as fractions of the maximum controller performance. The resulting workspace is shown in green in Figure. 1A.

F. Simulations

1) Control of Fatigable Time-Varying Models: In this simulation, we measured controller performance and muscle fatigue in musculoskeletal models representing an average male subject, a person with SCI, and a person with SCI after FES exercise, as described in sections II-B and II-C. A fourth condition, included as a control, used a musculoskeletal model with no fatigue or atrophy, as described in previous works [18], [19], [23], [24]. Here, the goal was to assess the feasibility of using existing DRL controller architectures for FES systems in individuals with chronic SCIs.

2) Assessing the Impact of Muscle Atrophy and Muscle Strength Asymmetry: We noticed that the DRL algorithm had difficulty converging on successful controllers if the flexor and extensor muscles fatigued in an asymmetrical manner. This could be due to the fact that (1) muscle forces after fatigue were not sufficient to move the arm to the target location and/or (2) the RL algorithm had difficulty with directionally asymmetrical actuator strengths. To test these hypotheses, we implemented models with non-fatigable (time-invariant) muscles, but with varying levels of maximum muscle forces. We tested 3 conditions:

- Symmetrical muscle strength: all maximum muscle forces were scaled by the same factor.
- Flexor force decrease: flexor muscles were weaker than extensor muscles, and all maximum flexor muscle forces were scaled by the same factor.
- Extensor force decrease: extensor muscles were weaker than flexor muscles, and all maximum extensor muscle forces were scaled by the same factor.

3) Controller Performance With Symmetrical Time-Varying Muscle Strengths: To demonstrate that muscle strength asymmetries can impair controller performance, we controlled for muscle strength asymmetries by artificially linking muscle fatigue levels between opposing muscles, as illustrated in Figure. 1A. Since flexor muscles displayed higher levels of fatigue than extensor muscles in preliminary simulations, we chose to have extensor fatigue track flexor fatigue. We used the same musculoskeletal models described in Simulation 1.

4) Rest-Reach Controller Training Protocol: In this simulation, we implemented a controller training protocol that included rests of 3 to 6 seconds between reaches to reduce overall levels of fatigue. We assessed the impact of rest duration on controller training times and controller performance. We evaluated the rest-reach protocol using the *SCI Exercised* model described above.

G. Controller Evaluation

For each condition, the controller training process shown in Figure. 2 was repeated 32 times, as was shown to provide an accurate assessment of controller performance in previous studies implementing the same motor task [23], [24]. Controller performance during training was measured every 5 minutes of simulated time for all simulations, except for the simulation where opposing muscles were artificially kept at the same levels of fatigue. For that simulation, we performed evaluations every minute to allow for a more thorough assessment of transient changes in controller performance due to time-varying muscle forces. During evaluations, DNN parameters were held constant while the controller performed 100 reaches. The success rate for an evaluation of a single controller was calculated as the number of successful reaches divided by the total number of reaches. The success rate for a condition was estimated as the median success rate of all controllers trained in that condition. Controller performance plots display the median success rate and the interquartile range during controller training. Training time was defined as the time taken for success rates to reach the maximum median performance for each condition. In workspace performance plots, evaluations included 2,000 reaches. The workspace was discretized into pixels of 0.3 × 0.3 cm, and performance was measured as the number of successful reaches divided by the total number of targets spawned within each pixel.

III. Results

A. DRL Control of Fatigable Musculoskeletal Models

Figure. 3A shows controller success rates as a function of training time for fatigable musculoskeletal models, in addition

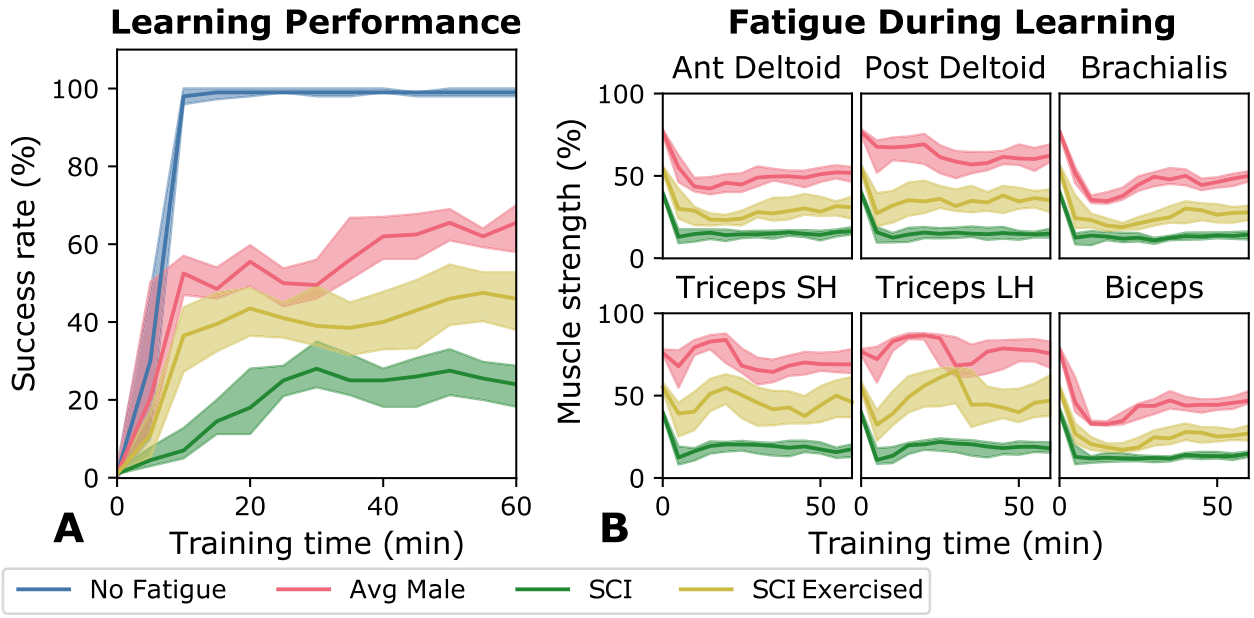

Fig. 3. **Controller training for fatigable musculoskeletal models. (A):** Controller success rates (successful reaches/total reaches) as a function of controller training time for different biomechanical models. **(B):** Effect of fatigue on relative muscle strength during controller training. Muscle strength decreases rapidly at the beginning of training for all conditions because of muscle fatigue.

TABLE II
DRL CONTROLLER PERFORMANCE (NO REST)

Model	Success Rate After 60 Minutes(%)	Training Time (min)
No Fatigue	99	15
Avg. Male	65	50
SCI Exercised	46	55
SCI Chronic	23	50

to a non-fatigable model that was included as a control. Success rates decreased substantially for fatigable models (Avg. Male, SCI, SCI Exercised), particularly for models with higher levels of atrophy, lower levels of muscle recovery (R), and higher levels of muscle fatigability (F). Success rates were lowest for the SCI model, which has the highest level of muscle atrophy and muscle fatigability, and the lowest level of muscle recovery. Training times, defined as the time taken for success rates to reach the maximum median performance, increased for each of the three models that included the effects of fatigue. Table II summarizes the success rates after 60 minutes of training for all conditions, as well as the training time for each condition.

Figure. 3B shows changes in muscle strength during controller training for each of the three models that include the effects of fatigue. Muscle strengths decreased rapidly due to fatigue at the beginning of training and plateaued thereafter.

B. DRL Control of Non-Fatigable Models Representing Atrophy

Figure. 4 shows success rates as a function of training time for different levels of muscle strength in a musculoskeletal model that does not include the effects of fatigue. In Figure. 4A, all muscle forces were scaled by the same factor, whereas in Figure. 4B, only the extensor muscles were weakened, and in Figure. 4C, only the flexor muscles were weakened. In Figures. 4A and 4B, the DRL controllers achieve success rates above 95% for all conditions representing atrophy levels of up to 75%. In Figure. 4C, the DRL controllers achieve success rates above 90% only when flexor atrophy was below 15%. An imbalance where the flexors were 30% weaker than the extensors (purple curve in Fig. 4C) caused success rates in this task to drop to approximately 67%. No targets were acquired when flexor forces were decreased by 90% or more.

Figure. 5 shows that impairments in controller performance were spatially isolated depending on the pattern of muscle atrophy. Symmetrical decreases in muscle forces caused the workspace to shrink symmetrically from the lateral regions towards the medial region of the workspace. Flexor-only atrophy caused the workspace to shrink from the contralateral region of the workspace towards the ipsilateral region of the workspace. Extensor-only atrophy caused the workspace to shrink from the ipsilateral region of the workspace towards the contralateral region of the workspace. Workspace decreases were most pronounced for the flexor-only atrophy condition.

C. DRL Performance for Symmetrical, Fatigable Musculoskeletal Models

Figure. 6A shows success rates during controller training for fatigable musculoskeletal models where muscle strengths from opposing muscles were artificially forced to be equal. When muscle strengths for opposing muscles were forced to be equal, controller performance was improved compared to models where opposing muscle strengths were allowed to be asymmetrical (compare Figure. 6A to Figure. 3A).

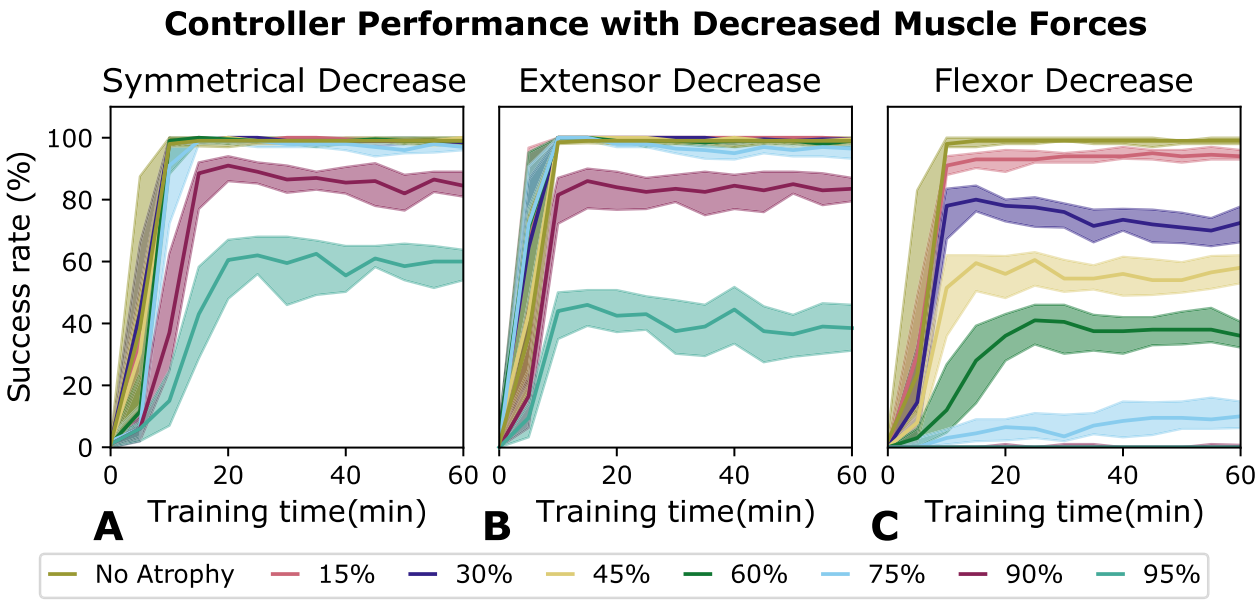

Fig. 4. **DRL controller training for non-fatigable muskuloskeletal models representing atrophy. (A):** Muscle strength was scaled by the same factor for all muscles. Note that the DRL controller requires only very low levels of muscle strength to operate effectively. DRL controllers perform well when muscle atrophy is symmetrical. **(B):** Muscle strength was decreased only for the extensor muscles. Controller performance remained high, even for high levels of muscle atrophy. **(C):** Muscle strength was decreased only for the flexor muscles. The controller was impaired by asymmetries where the flexor muscles were weaker than the extensors. Success rates for flexor atrophy levels above 75% were nearly zero - notice that the learning curves representing 90% and 95% atrophy coincide with the x-axis. Taken together, these results demonstrate that muscle strength asymmetries can considerably impair controller performance, even in the absence of muscle fatigue during training.

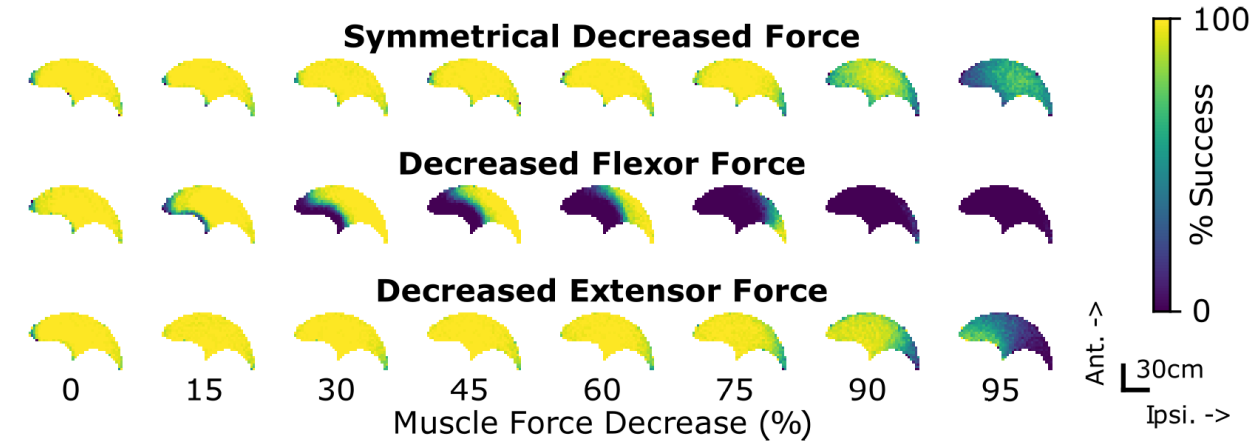

Fig. 5. **DRL controller performance throughout the workspace for non-fatigable models of atrophy.** Each plot displays the unfatigued workspace of the arm, with the colored shading indicating the success rate at each location within that workspace. Symmetrical decreases in muscle strengths cause the workspace to shrink from the edges of the workspace towards the center as forces decrease. Asymmetrical decreases in muscle strengths cause the workspace to shrink asymmetrically from the corresponding sides of the workspace.

For example, the success rates for the *Avg. Male* model after 60 minutes of training went from 65% for asymmetrical models to almost 100% for symmetrical models. For the *SCI Exercised* model, success rates after 60 minutes of training went from 46% to approximately 90%. For the model representing people with chronic SCI and no FES exercise, the increase in success rates was less substantial - improving from 23% to 33%.

D. Rest-Reach Training Protocol

Figure. 7 shows success rates and fatigue levels as functions of training time while using a controller training paradigm that included rest periods between reaches. The musculoskeletal model used in this simulation represented a subject with SCI after FES exercise. Success rates were approximately 95% after 120 minutes of training for rest periods of 5 seconds after every reach. After introducing a pause of at least 3 seconds between reaches, muscle strengths remained above 70% for all muscles included in this model.

IV. Discussion

People with chronic paralysis have muscles that fatigue more quickly [11] as well as high levels of disuse atrophy [13]. In this work, we extended existing musculoskeletal models

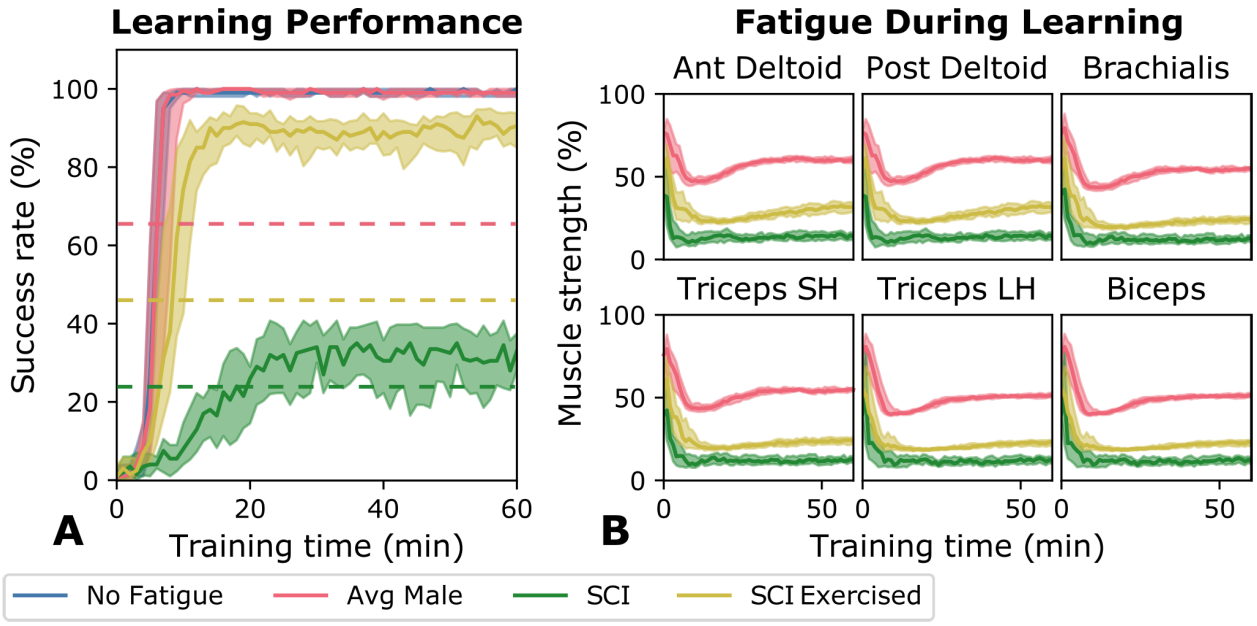

Fig. 6. **DRL controller training for fatigable musculoskeletal models where muscle strengths were artificially linked between opposing muscles. (A):** Controller performance as a function of controller training time. Controllers are more successful when used for models with symmetrical fatigue compared to models with asymmetrical fatigue (Figure 3). For comparison, the dashed lines represent the success rates after 60 minutes of training in Figure 3. Note that controller performance for the SCI model remains poor, even when muscle strengths are symmetrical, because of the combined negative effects of high levels of muscle atrophy and high levels of muscle fatigue. **(B):** Muscle fatigue during training. Fatigue levels were linked between antagonistic muscle pairs to eliminate the effects of asymmetry, specifically the anterior deltoid and the posterior deltoid, the brachialis and the short head of the triceps, and the long head of the triceps and the biceps.

to simulate these conditions and generated more realistic predictions of DRL controller performance for people with SCIs. Here, we demonstrated that DRL controllers work well under functionally realistic conditions (exercised SCI muscles), as long as short rests are provided in between reaches. Interestingly, we uncovered that the major challenge for DRL controllers was not decreased muscle strength in itself, but rather muscle strength asymmetries.

A. Impact of Muscle Atrophy and Muscle Strength Asymmetries on DRL Controller Performance

The simulations in the current study suggested that DRL controllers can effectively coordinate muscle activations to perform reaches in a large workspace, even with low maximum muscle forces due to muscle atrophy. However, very high rates of atrophy ($> 70\%$) could impair controller performance if combined with high levels of muscle fatigue. Also, these simulations revealed that large asymmetries ($> 30\%$) where flexor muscles are weaker than extensor muscles can impair DRL control for the task studied in this work.

To the best of our knowledge, these are the first simulations that characterize the impact of muscle strength asymmetries on RL control of biomechanical systems. A previous study [19] obtained good results while using RL to control a biomechanical model where flexor muscles were 50% weaker than extensor muscles. However, the workspace used in [19] was contained in a region where controllers in this work acquired more than 90% of the targets for flexor atrophy levels of 45%. The current study expanded on [19], not only by increasing the size of the workspace, but also by characterizing controller performance at various levels of asymmetrical atrophy. This parameter sweep revealed that not all patterns of muscle strength asymmetries impair controller performance. For this arm model, flexor-only muscle atrophy impaired controller performance far more than similar levels of extensor-only atrophy, possibly because the workspace extended farther in the direction of joint flexion.

Since the arm model was supported against gravity, we encourage caution when predicting results for other tasks. Nonetheless, these results are relevant to people with SCIs who use mobile arm supports [10], [37].

B. Impact of Muscle Strength Time-Variance Due to Fatigue on DRL Controller Performance

The results from Figure. 6A suggest that muscle strength time-variance due to muscle fatigue is unlikely to impair DRL control if the remaining muscle strength is sufficient to perform the task. Note that for the *Chronic SCI* model, combined levels of atrophy and fatigue cause muscle strength to drop below 4% of the muscle strength available to non-disabled people. Considering the results in Figure. 4A, 4% of non-disabled muscle strength is likely to be insufficient to perform the task. This would explain why the *Chronic SCI* condition, compared to other conditions, did not benefit as much from forcing fatigue levels in opposing muscles to be equal. These results highlight the importance of adequate FES-exercise to increase muscle strength in chronically paralyzed muscles prior to controller deployment.

As discussed above, fatigue-induced performance decreases were likely caused in large part by muscle strength

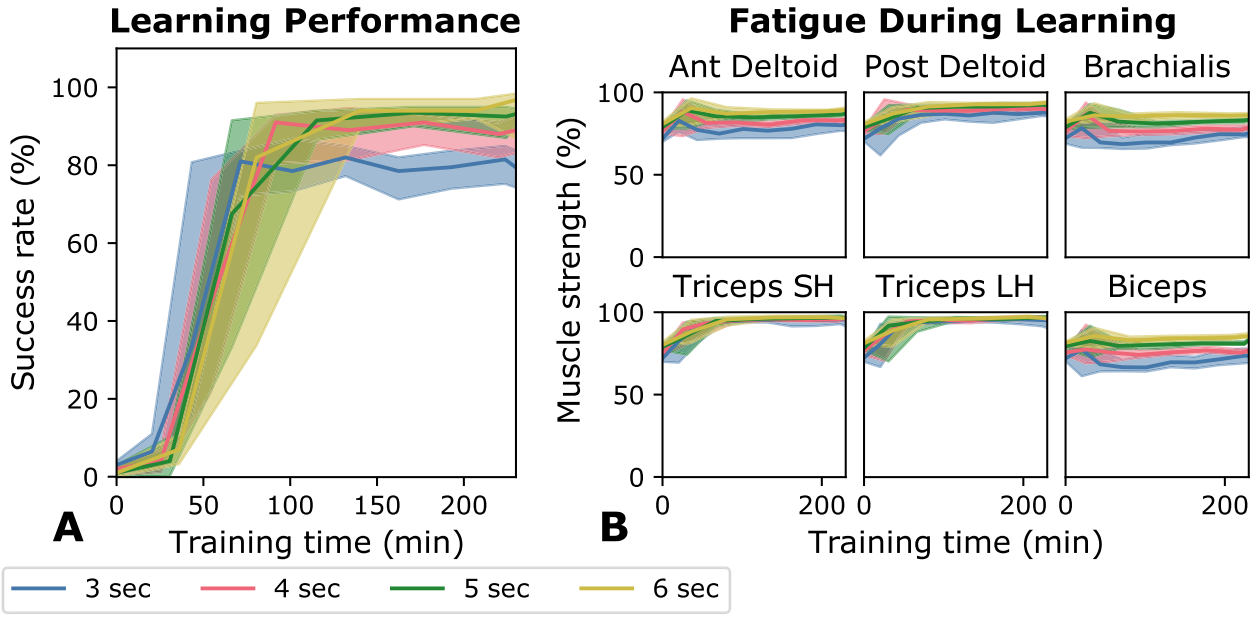

Fig. 7. **DRL controller training with rest between reaches for the SCI Exercised model. (A):** Controller performance as a function of controller training time. Allowing for rest between reaches considerably improves controller success rates. For example, a pause of 5 seconds between reaches results in success rates of around 90% for the studied motor task. **(B):** Fatigue levels during training. Introducing short pauses in between reaches allows muscle strengths to remain high, and nearly constant, throughout controller use.

asymmetries, rather than by muscle strength insufficiency. The impact of muscle strength asymmetries caused by fatigue was substantially ameliorated by introducing rests in between reaches (Figure. 7). For the motor task studied in this work, a 5 second rest was sufficient to limit muscle fatigue for a model representing people with SCI after FES-exercise, and it resulted in success rates above 95%.

In clinical practice, it is unlikely that users would continuously move their arms across the workspace. Normal upper-limb activity includes frequent periods of inactivity. Therefore, we expect that sufficient muscle recovery should be achieved in naturally-occurring periods of inactivity. During controller training, rest duration in between reaches can be optimized considering patient biomechanics and task complexity.

For non-fatigable models, training was completed in less than 20 minutes. When adding rest periods to avoid fatigue-induced muscle strength asymmetries, training time remained below 120 minutes. While this is a substantial increase in training time, training for 120 minutes is still practical. Additionally, training times might be decreased by using techniques such as transfer learning [24].

C. Motor Behavior and Controller Strategy

In this study, as in previous implementations of DRL controllers [23], [24], we observed that commanded muscle activations were frequently either below 5% or above 95%. Supplementary Figure. 1 shows a histogram of commanded muscle activations for the simulation in Figure. 3 after controller training. All other conditions implemented in this work displayed the same trend. An "all-or-nothing" controller strategy may have exacerbated the development of high levels of fatigue because fatigue increases exponentially with increasing muscle activation. Interestingly, increasing the penalty on muscle activations did not result in a different controller strategy. Higher muscle activation penalties caused minor shifts in the muscle activation histogram, and they resulted in lower success rates, as can be seen in Supplementary Figures. 2 and 3.

Qualitatively, the motor behavior observed for successfully trained controllers could be described as smooth and fast with deceleration as the endpoint of the arm approached the target region. The arm was generally stable at the target region for all conditions (see Supplementary Figure. 7). The duration of the reaches for the fatigable models was approximately 0.4 seconds, which is slightly longer than the duration reported in a similar study [23]. Yet, this is not surprising, since the workspace used in this study was considerably larger, and muscle strength was smaller due to the inclusion of atrophy and fatigue.

D. DRL Controller Use in Patients With SCI

Chronic paralysis causes muscle atrophy [13], and more rapid muscle fatigue [11]. Additionally, SCI is likely to cause denervation in certain muscles [13], and FES exercise cannot reverse atrophy caused by denervation. Furthermore, muscle strength asymmetries between opposing muscles are common even for non-disabled individuals [41]. Therefore, patients with SCIs are likely to exhibit muscle strength asymmetries, and it is possible that these asymmetries are characterized by large imbalances where flexors are weaker than extensors. Since these results indicate that such imbalances impair DRL controller performance, it may be necessary to scale electrode stimulation ranges to compensate for flexor weakness.

Figure. 5 shows that decreases in controller success rates due to muscle atrophy are not evenly distributed across

the workspace. Controller performance is impaired mostly at the lateral borders of the workspace, and the affected area increases towards the center of the workspace with increasing levels of atrophy. Even for the most challenging condition, where flexors were considerably weaker than extensors, a large portion of the workspace remained reachable. Therefore, muscle strength asymmetries may decrease the reachable workspace, but they do not seem to substantially impair controller performance within the reachable workspace. Taken together, these findings suggest that the task workspace could be optimized by considering the anatomy of the patient.

The results in this study (see Figures. 6A and 7A) indicate that time-variance due to muscle fatigue is unlikely to affect controller performance in reaching tasks with the arm supported against gravity, as long as rest periods are included between reaches. However, we evaluated DRL controller performance in a motor task that did not require precise movements. More complex motor tasks requiring fine motor control may be affected by muscle strength time-variance.

High levels of muscle activation, which occur frequently when using the control strategy implemented by this DRL algorithm, could impair fine motor control, in addition to being potentially unsafe and uncomfortable for patients. Therefore, we recommend that the electrodes are profiled in order to set safe and comfortable levels of stimulation before controller training. Profiling electrodes to ensure safety is a standard practice in neuroprosthesis implementation [10], [37].

We highlight the importance of exercising chronically paralyzed muscles using FES to partially reverse disuse atrophy prior to controller deployment. Also, controller use should include frequent periods of inactivity. When high levels of muscle atrophy were combined with muscle fatigue, the resulting muscle strengths were not sufficient to perform the relatively simple task studied in this work. However, since upper-limb use naturally includes periods of low activity, we do not anticipate that this will affect the use of the controllers, or cause inconvenience for the users.

Theoretically, the proposed RL controller can adapt to the real-world properties of persons with SCI, which is a primary benefit of this model-free approach. Therefore, the use of a musculoskeletal model is not required. However, a subject-specific musculoskeletal model could be used to pre-train the controller and thus reduce the training time.

E. Limitations

In this work, we used a musculoskeletal model where the arm was supported against gravity. The model only included 6 simple Hill-type actuators, and the model did not include joint stiffness or muscle spasticity, which are common in patients with SCI. Also, the motor task studied in this work was relatively simple, and it involved 2 DoFs. These were deliberate study design choices, as in previous works investigating the early feasibility of DRL for motor control [18], [19], [23], [24], [42], [43], [44], to limit the impact of confounding variables. Here, the goal was to characterize the impact of fatigue and time-varying muscle strength on DRL controllers. Future studies should implement DRL control of more complex musculoskeletal models that can more accurately represent the biomechanics of people with chronic paralysis, as well as more meaningful functional tasks. A recent study demonstrated that similar RL controllers were robust to changes in many biomechanical parameters [24], [45], suggesting that these controllers can generalize across different users with minimal retraining. However, we anticipate that motor tasks involving object interaction will require major adjustments to the reward function.

We used data from the literature to estimate levels of atrophy, as well as the R and F coefficients for the musculoskeletal models in this work. However, these data, particularly the fatigue curves used to estimate the R and F coefficients, were recorded during motor tasks that differed in important ways. Specifically, suitable data to extract R and F coefficients was available either during MVC [34], or intermittent tasks involving full activation and complete rest [12]. Yet, previous studies argue that R and F coefficients are likely to change depending on the level of muscle contraction, due to physiological factors such as increased blood flow in muscles that are relaxed [33]. Therefore, we anticipate that more accurate estimations of R and F coefficients could lead to shifts in the expected success rates (see Supplementary Figures. 4 and 5) and fatigue levels for the conditions studied in this work, although we expect that the main conclusions would remain unchanged.

Finally, the R and F coefficients representing people with SCIs were estimated using soleus data. The non-paralyzed soleus is substantially less fatigable than upper-limb muscles due to its higher proportion of type I muscle fibers [14]. However, chronically-paralyzed muscles become primarily composed of type II fibers [14], which fatigue more quickly. Therefore, we assumed that different muscle groups displayed similar fatigability after SCI. Yet, it is possible that small differences in muscle composition remain after paralysis-driven muscle adaptation.

V. Conclusion

In this study, we have implemented a musculoskeletal model of the arm that allows for continuous estimation of muscle fatigue during motor tasks. We have adapted this model to represent patients with SCI before and after FES-exercise, and we have used this model to investigate the feasibility of DRL control of FES for people with chronic paralysis. The results suggest that DRL controllers should provide good performance if muscles are exercised to reverse disuse atrophy, and if there are realistic rest periods between movements. However, the simulations indicate that large muscle strength asymmetries may impair DRL controllers. Muscle strength asymmetries that occur due to fatigue may be alleviated by introducing frequent rests during controller use. We recommend that electrodes are carefully profiled and that stimulation ranges are scaled to compensate for large asymmetries due to disuse atrophy or denervation, and to ensure that movements are safe and comfortable for the users. These results support the feasibility of using DRL to control FES systems to reverse paralysis due to SCI.

ACKNOWLEDGMENT

This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University.

REFERENCES

[1] International Campaign for Cures of Spinal Cord Injury. (2011). *General Information*. [Online]. Available: https://campaignforcure.org/index.php/general-information

[2] *Spinal Cord Injury Facts and Figures at a Glance*, Nat. Spinal Cord Injury Stat. Center, Birmingham, AL, USA, 2021, pp. 1–2.

[3] C. Donnelly *et al.*, "Client-centred assessment and the identification of meaningful treatment goals for individuals with a spinal cord injury," *Spinal Cord*, vol. 42, no. 5, pp. 302–307, May 2004.

[4] K. D. Anderson, "Targeting recovery: Priorities of the spinal cord-injured population," *J. Neurotrauma*, vol. 21, no. 10, pp. 1371–1383, Oct. 2004.

[5] J. L. Collinger *et al.*, "Functional priorities, assistive technology, and brain-computer interfaces after spinal cord injury," *J. Rehabil. Res. Dev.*, vol. 50, no. 2, pp. 145–160, 2013.

[6] M. W. Keith *et al.*, "Implantable functional neuromuscular stimulation in the tetraplegic hand," *J. Hand Surg.*, vol. 14, no. 3, pp. 524–530, May 1989.

[7] K. L. Kilgore *et al.*, "An implanted upper-extremity neuroprosthesis. Follow-up of five patients," *J. Bone Joint. Surg.*, vol. 79, no. 4, pp. 533–541, 1997.

[8] P. H. Peckham *et al.*, "Efficacy of an implanted neuroprosthesis for restoring hand grasp in tetraplegia: A multicenter study," *Arch. Phys. Med. Rehabil.*, vol. 82, no. 10, pp. 1380–1388, Oct. 2001.

[9] M. R. Popovic, D. B. Popovic, and T. Keller, "Neuroprostheses for grasping," *Neurol. Res.*, vol. 24, no. 5, pp. 443–452, 2002.

[10] W. D. Memberg *et al.*, "Implanted neuroprosthesis for restoring arm and hand function in people with high level tetraplegia," *Arch. Phys. Med. Rehabil.*, vol. 95, no. 6, pp. 1201–1211, 2014.

[11] M. Barat, P. Dehail, and M. de Seze, "Fatigue after spinal cord injury," *Ann. Phys. Rehabil. Med.*, vol. 49, no. 6, pp. 365–369, 2006.

[12] R. K. Shields, "Fatigability, relaxation properties, and electromyographic responses of the human paralyzed soleus muscle," *J. Neurophysiol.*, vol. 73, no. 6, pp. 2195–2206, Jun. 1995.

[13] C. K. Thomas, E. Y. Zaidner, B. Calancie, J. G. Broton, and B. R. Bigland-Ritchie, "Muscle weakness, paralysis, and atrophy after human cervical spinal cord injury," *Exp. Neurol.*, vol. 148, no. 2, pp. 414–423, 1997.

[14] B. Biering-Sørensen, I. B. Kristensen, M. Kjær, and F. Biering-Sørensen, "Muscle after spinal cord injury," *Muscle Nerve*, vol. 40, no. 4, pp. 499–519, 2009.

[15] D. Blana, R. F. Kirsch, and E. K. Chadwick, "Combined feedforward and feedback control of a redundant, nonlinear, dynamic musculoskeletal system," *Med. Biol. Eng. Comput.*, vol. 47, no. 5, pp. 533–542, May 2009.

[16] M. O. Ibitoye, N. A. Hamzaid, N. Hasnan, A. K. A. Wahab, and G. M. Davis, "Strategies for rapid muscle fatigue reduction during FES exercise in individuals with spinal cord injury: A systematic review," *PLoS ONE*, vol. 11, no. 2, pp. 1–28, 2016.

[17] A. J. Buckmire, T. J. Arakeri, J. P. Reinhard, and A. J. Fuglevand, "Mitigation of excessive fatigue associated with functional electrical stimulation," *J. Neural Eng.*, vol. 15, no. 6, pp. 1–21, 2018.

[18] K. M. Jagodnik, P. S. Thomas, A. J. van den Bogert, M. S. Branicky, and R. F. Kirsch, "Human-like rewards to train a reinforcement learning controller for planar arm movement," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 5, pp. 723–733, May 2016.

[19] K. M. Jagodnik, P. S. Thomas, A. J. van den Bogert, M. S. Branicky, and R. F. Kirsch, "Training an actor-critic reinforcement learning controller for arm movement using human-generated rewards," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1892–1905, Oct. 2017.

[20] D. D. Febbo *et al.*, "Does reinforcement learning outperform PID in the control of FES-induced elbow flex-extension?" in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2018, pp. 1–6.

[21] I. Akkaya *et al.*, "Solving Rubik's cube with a robot hand," 2019, *arXiv:1910.07113*.

[22] M. Andrychowicz *et al.*, "Hindsight experience replay," in *Proc. Adv. Neural. Inf. Process. Syst. (NIPS)*, Dec. 2017, pp. 5049–5059.

[23] D. C. Crowder, J. Abreu, and R. F. Kirsch, "Hindsight experience replay improves reinforcement learning for control of a MIMO musculoskeletal model of the human arm," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1016–1025, 2021.

[24] D. C. Crowder, J. Abreu, and R. F. Kirsch, "Improving the learning rate, accuracy, and workspace of reinforcement learning controllers for a musculoskeletal model of the human arm," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 30–39, 2022.

[25] K. T. Gelenitis, B. M. Sanner, R. J. Triolo, and D. J. Tyler, "Selective nerve cuff stimulation strategies for prolonging muscle output," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 5, pp. 1397–1408, May 2020.

[26] K. Gelenitis, M. Freeberg, and R. Triolo, "Sum of phase-shifted sinusoids stimulation prolongs paralyzed muscle output," *J. NeuroEng. Rehabil.*, vol. 17, no. 1, pp. 1–7, Dec. 2020.

[27] T. Xia and L. A. F. Law, "A theoretical approach for modeling peripheral muscle fatigue and recovery," *J. Biomech.*, vol. 41, no. 14, pp. 3046–3052, Oct. 2008.

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

[29] B. A. Garner and M. G. Pandy, "Estimation of musculotendon properties in the human upper limb," *Ann. Biomed. Eng.*, vol. 31, no. 2, pp. 207–220, Feb. 2003.

[30] N. Bhushan and R. Shadmehr, "Computational nature of human adaptive control during learning of reaching movements in force fields," *Biol. Cybern.*, vol. 81, no. 1, pp. 39–60, Jul. 1999.

[31] D. A. Winter, *Biomechanics and Motor Control of Human Movement*, 4th ed. Hoboken, NJ, USA: Wiley, 2009.

[32] L. A. Frey-Law, J. M. Looft, and J. Heitsman, "A three-compartment muscle fatigue model accurately predicts joint-specific maximum endurance times for sustained isometric tasks," *J. Biomech.*, vol. 45, no. 10, pp. 1803–1808, Jun. 2012.

[33] J. M. Looft, N. Herkert, and L. Frey-Law, "Modification of a three-compartment muscle fatigue model to predict peak torque decline during intermittent tasks," *J. Biomech.*, vol. 77, pp. 16–25, Aug. 2018.

[34] J. Z. Liu, R. W. Brown, and G. H. Yue, "A dynamical model of muscle activation, fatigue, and recovery," *Biophys. J.*, vol. 82, no. 5, pp. 2344–2359, May 2002.

[35] R. K. Shields and S. Dudley-Javoroski, "Musculoskeletal plasticity after acute spinal cord injury: Effects of long-term neuromuscular electrical stimulation training," *J. Neurophysiol.*, vol. 95, no. 4, pp. 2380–2390, 2006.

[36] D. Sale, J. Quinlan, E. Marsh, A. McComas, and A. Belanger, "Influence of joint position on ankle plantarflexion in humans," *J. Appl. Physiol.*, vol. 52, no. 6, pp. 1636–1642, 1982.

[37] A. B. Ajiboye *et al.*, "Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: A proof-of-concept demonstration," *Lancet*, vol. 389, no. 10081, pp. 1821–1830, 2017.

[38] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[39] S. Fujimoto, H. Van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, vol. 4, 2018, pp. 2587–2601.

[40] A. Raffin *et al.*, "Stable-baselines3: Reliable reinforcement learning implementations," *J. Mach. Learn. Res.*, vol. 22, no. 268, pp. 1–8, 2021.

[41] V. Serrau, T. Driss, H. Vandewalle, D. G. Behm, E. Lesne-Chabran, and A. L. Pellec-M'uller, "Muscle activation of the elbow flexor and extensor muscles during self-resistance exercises," *J. Strength Conditioning Res.*, vol. 26, no. 9, pp. 2468–2477, 2012.

[42] T. Ehsan, J. A. Homayoun, and F. Ali, "Learning to control the three-link musculoskeletal arm using actor–critic reinforcement learning algorithm during reaching movement," *Biomed. Eng.*, vol. 26, no. 5, 2014, Art. no. 1450064.

[43] G. Vahid, L. Caro, and P. Mohamad, "Application of actor-critic reinforcement learning method for control of a sagittal arm during oscillatory movement," *Biomed. Eng.*, vol. 16, no. 6, pp. 305–312, 2004.

[44] V. Golkhou, M. Parnianpour, and C. Lucas, "The role of multisensor data fusion in neuromuscular control of a sagittal arm with a pair of muscles using actor-critic reinforcement learning method," *Technol. Health Care*, vol. 12, no. 6, pp. 425–438, Jan. 2005.

[45] D. C. Crowder, "Reinforcement learning for control of a multi-input, multi-output model of the human arm," Ph.D. dissertation, Dept. Biomed. Eng., Case Western Reserve Univ., Cleveland, OH, USA, 2021.