NORTHWESTERN UNIVERSITY

Compact Deep Learning Models and Auxiliary Methods for Robust Myoelectric Control

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Biomedical Engineering

By Yuni Teh

EVANSTON, ILLINOIS

June 2022

Abstract

Compact Deep Learning Models and Auxiliary Methods for Robust Myoelectric Control

Yuni Teh

Myoelectric pattern recognition-based upper limb prostheses measure electromyographic (EMG) signals from the residual limb and learn to identify muscle activity patterns that correspond to intended gestures. To train an accurate pattern recognition controller, it is essential that the training signals typify signals measured in real-world scenarios. When these conditions are met, clinical systems enable accurate and intuitive prosthesis control. However, routine usage of a prosthesis gives rise to signal nonstationarities that cause dataset shifts (ie. changes in the joint distribution of classifier input and output). These shifts reduce classification accuracy and render control ineffective. In this dissertation, I examine common sources of dataset shift that affect myoelectric pattern recognition and propose clinically feasible approaches to improve control robustness. First, I investigate the effects of limb position and external load on real-time pattern recognition control and show that a modified training data collection protocol can eliminate these effects in amputee users. Next, I combine data augmentation and deep learning techniques to build classifiers that are tolerant to multi-channel signal noise originating from the electrode-skin interface. Finally, I quantify dataset shift across long-term prosthesis usage and use continual learning with deep neural networks to reduce classifier recalibration frequency. Together, these methods provide a foundation for clinical implementations of advanced deep learning controllers that are robust under dataset shift.

Acknowledgments

Thanks to Levi for always making time, for giving me the freedom and tools to find my way, and for his confidence in me. To my committee, Eric Perreault, Matthew Tresch, and Brenna Argall for their encouragement and advice.

Thanks to the Northwestern BME department, Shirley Ryan AbilityLab, and the Center for Bionic Medicine. To the CBM scholars, past and present, for the collaboration and commiseration. To Kunal, Chandler, Maggie for sharing snacks and memes.

Thanks to the people who helped me keep it together: Myra, Ching, Katrina, Stephanie, Max, Drew, Alex, Joseph, Hannah.

Thanks to my family for their unconditional support.

Table of Contents

List of Figures and Tables

1 Introduction

1.1 Motivation

In the United States, there are more than 41,000 people who live with an upper limb amputation as a result of trauma, dysvascular disease, or cancer, 40% of whom are transradial amputees (Dillingham et al., 2002; Ziegler-Graham et al., 2008). Individuals with major upper limb amputations may have difficulty performing essential activities of daily living, such as eating, grooming, and dressing (Datta et al., 2004; Davidson, 2002; Jang et al., 2011; Routhier et al., 2001). In many cases, an amputation also affects an individual's capacity to return to employment and causes a decline in overall quality of life (Fernandez et al., 2000) .

The use of a prosthesis can restore some functionality and independence; 91% of amputees that returned to work used a prosthesis (Fernandez et al., 2000). While some people opt for cosmetic prostheses for aesthetic reasons, others choose to use powered prostheses. These may include body-powered devices or electric devices powered by motors. Body-powered devices may be more durable, while electric devices may provide greater dexterity and precision.

Myoelectric control measures electrical signals from muscle contractions and uses them as control signals, offering an intuitive method to control electric prostheses. Commercial devices employ two main types of control systems: amplitude-based direct control and pattern recognition (PR) control. Direct control methods use agonist-antagonist muscle pairs to control one degree-of-freedom at a time (Williams et al., 2004). Conversely, pattern recognition control measures an array of muscle signals and identifies patterns of muscle activity that correspond to intended gestures.

Although myoelectric devices have been around for decades, user adoption is still low, with many citing a lack of reliable control method or inconvenience as reasons (Biddiss & Chau, 2007a). In 2007, the rejection rate of myoelectric hands was 39% (Biddiss & Chau, 2007b). In 2020, the rejection rate was 44%, suggesting that despite the advances over the past decade, amputee users are still not satisfied with current technology (Salminger et al., 2020).

1.2 Background

Myoelectric pattern recognition algorithms

Myoelectric pattern recognition control for upper limb prostheses dates back to the 1960s but has only recently been made clinically available (Whitney, 1969). With the adoption of this technology, users have seen functional improvements (L. J. Hargrove et al., 2013). Still, the inner workings of clinically available control algorithms are predominantly based on techniques introduced almost thirty years ago, leaving much room for further advancements (Hudgins et al., 1993).

PR control methods measure electromyographic (EMG) signals from an array of electrodes and learn the patterns of muscle activity that correspond to intended movements (Englehart & Hudgins, 2003; Hudgins et al., 1993; Parker et al., 2006). Typically, PR controllers use a classifier to classify time domain and autoregressive features extracted from windowed EMG signals. PR control is effective, provided the EMG interface is stable and the signals are repeatable (L. J. Hargrove et al., 2013).

Real-world factors that influence control performance

Robust PR control is an ongoing area of research. The versatility of upper limb functions creates nonstationary situations in which EMG signals can change. When there is a mismatch between the distribution of the training EMG signals and test EMG signals, PR control

performance declines. This distribution mismatch, termed dataset shift, underlies most of the factors that negatively influence control performance.

Dataset shift in EMG control can be caused by changes in generated signals and changes in the measured signals. For example, muscle activation differences due to limb position, load, or fatigue cause changes in the generated signals, shifting the test distribution away from the training distribution (Campbell, Phinyomark, et al., 2020; Anders Fougner et al., 2011). On the other hand, signal noise, electrode shift, sweat, and other disturbances at the electrode cause changes in the measured signals, which also shifts the test distribution away from the training distribution (Engdahl et al., 2015). Numerous methods have been proposed to reduce the effects of dataset shift with varying levels of success; nevertheless, it still poses a problem to current prosthesis users.

Deep learning-based myoelectric control

In recent years, the use of deep learning to enable myoelectric control has rapidly gained traction. Deep learning models combine nonlinear activation functions and modular network topologies to learn complex structures within data. Various models have been used to enable simultaneous control, incorporate computer vision, generate training signals, adapt controllers across users, and improve robustness to dataset shift (Ameri et al., 2019, 2020; Côté-Allard et al., 2019; Park & Lee, 2016; ur Rehman et al., 2018; D. Yang et al., 2021). Despite the growing popularity of deep learning, the computational requirements of these models hinder their clinical implementation.

1.3 Specific Aims

The overall objective of this dissertation is to develop and evaluate clinically feasible machine learning algorithms and protocols that can improve myoelectric pattern recognition

control robustness to common sources of dataset shift. To this end, I propose the following specific aims.

Aim 1. Quantify and compare the effects of limb position and external load on real-time pattern recognition control.

Studies have indicated that LDA-based prosthesis controllers are inaccurate when trained in one limb position and tested in another – an example of dataset shift induced by sample selection bias. However, investigations of the "limb position effect" in individuals with upper limb loss were either conducted offline or without a physical prosthesis on the residual limb, thus limiting the clinical applicability of their findings. There remains a gap in our knowledge of how limb position affects real-time control in amputee users when the residual limb is supporting the weight of a prosthesis. This aim seeks to understand the individual and interactive effects of limb position and external load and determine if they can be reduced by using an improved training data collection protocol.

Aim 2. Apply machine learning techniques to learn latent representations of muscle activation patterns that are robust to interface noise.

Changes at the electrode-skin interface (eg. electrode liftoff, broken wires, residual limb volume fluctuations) can generate signal disturbances that reduce classification accuracy. Previous studies proposed filters or adaptive algorithms to mitigate these effects. However, filtering methods are limited in their ability to generalize across intermittent and periodic noise and adaptation methods require additional processing steps. To our knowledge, there are no clinically feasible classifiers that are inherently robust to interface noise. The purpose of this aim is to develop and evaluate a noise-tolerant classifier that can learn and classify robust discriminative latent features shared by clean and noisy EMG signals.

Aim 3. Develop fast adaptation methods for deep convolutional neural networks and evaluate their effects on long-term myoelectric control stability and accuracy.

When dataset shift reduces pattern recognition accuracy and renders control ineffective, prosthesis users must recalibrate their devices to regain functionality. Clinically available control algorithms are not yet robust to common sources of dataset shift and therefore require frequent recalibrations. Although deep learning methods, such as those in Aim 2, may improve robustness to dataset shift, it is unrealistic at this stage to expect a control strategy that never has to be recalibrated or updated. However, effective recalibration or adaptation of a deep learning model nontrivial. Fully retraining a model is computationally and memory intensive whereas adapting a model may lead to catastrophic forgetting, or the loss of previously learned information. In this aim, I develop fast adaptation methods for deep convolutional neural network classifiers that can learn from new training data without catastrophic forgetting. I implement these methods on EMG data spanning 6 months to evaluate classification and recalibration performance over time.

1.4 Dissertation Overview

This dissertation is organized as follows. Aim 1 will be discussed in Chapter 2, which was published as an article in IEEE Transactions on Neural Systems and Engineering, Aim 2 will be discussed in Chapter 3, which is in review for Journal of NeuroEngineering and Rehabilitation, and Aim 3 will be discussed in Chapter 4. Finally, I summarize my contributions, their limitations, and potential future directions in Chapter 5.

2 Understanding the effects of limb position and external load on real-time pattern recognition control in amputees

2.1 Abstract

Limb position is a factor that negatively affects myoelectric pattern recognition classification accuracy. However, prior studies evaluating impact on real-time control for upper-limb amputees have done so without a physical prosthesis on the residual limb. It remains unclear how limb position affects real-time pattern recognition control in amputees when their residual limb is supporting various weights. We used a virtual reality target achievement control test to evaluate the effects of limb position and external load on real-time pattern recognition control in fourteen intact limb subjects and six major upper limb amputee subjects. We also investigated how these effects changed based on different control system training methods. In a static training method, subjects kept their unloaded arm by their side with the elbow bent whereas in the dynamic training method, subjects moved their arm throughout a workspace while supporting a load.

When static training was used, limb position significantly affected real-time control in all subjects. However, amputee subjects were still able to adequately complete tasks in all conditions, even in untrained limb positions. Moreover, increasing external loads decreased controller performance, albeit to a lesser extent in amputee subjects. The effects of limb position did not change as load increased, and vice versa. In intact limb subjects, dynamic training significantly reduced the limb position effect but did not completely remove them. In contrast, in amputee subjects, dynamic training eliminated the limb position effect in three out of four outcome measures. However, it did not reduce the effects of load for either subject population. These findings suggest that results obtained from intact limb subjects may not generalize to amputee

subjects and that advanced training methods can substantially improve controller robustness to different limb positions regardless of limb loading.

2.2 Introduction

In the United States, there are approximately 41,000 people living with a major upper limb amputation, 40% of whom are transradial amputees (Dillingham et al., 2002; Ziegler-Graham et al., 2008). These individuals have difficulty performing essential activities of daily living, such as eating, grooming, and dressing (Datta et al., 2004; Davidson, 2002; Jang et al., 2011; Routhier et al., 2001). In many cases, an amputation also affects an individual's capacity to return to employment and causes a decline in overall quality of life(Fernandez et al., 2000) .

Myoelectric prostheses have the potential to restore the function required to regain basic independence. Compared to body-powered prostheses, these devices provide more degrees of freedom (DOF's), allowing for finer control and potentially fewer compensatory movements(Montagnani et al., 2015) . Nevertheless, despite decades of advancements, challenges persist in controlling myoelectric prostheses. This remains an active area of research and several methods have been proposed to improve control.

One such method is myoelectric pattern recognition (PR), which relies on separable and repeatable muscle contractions. PR controllers have been validated extensively in lab settings and are now commercially available; however, several factors impact their real-world performance. Limb position has often been reported as one of these factors. Many studies have shown that offline classification errors increase significantly when different limb positions are tested (Campbell, Phinyomark, et al., 2020; Castellini et al., 2009; D. et al., 2017; Anders Fougner et al., 2011; Khushaba et al., 2014; L. Chen, Y. Geng, 2011; Liu et al., 2014). Several groups have quantified

the changes in detected signals that underly these poor classification rates (Betthauser et al., 2018; Boschmann & Platzner, 2014; A. Radmand et al., 2014). These results have prompted the development of new training and control methods to mitigate the "limb position effect"(Betthauser et al., 2018; Anders Fougner et al., 2011; Y. Geng, Zhang, et al., 2012; Hwang et al., 2017; Khushaba et al., 2014; Ashkan Radmand et al., 2014; Woodward & Hargrove, 2019; D. Yang et al., 2017) .

Most of these experiments only evaluated offline performance, which does not always correlate with real-time controllability (L. Hargrove et al., 2007; Ortiz-Catalan et al., 2015). To our knowledge, only two studies evaluated the real-time effects of limb position on PR control in amputees (Y. Geng, Zhou, et al., 2012; Woodward & Hargrove, 2019). These effects were found to be minor; however, these studies were limited because amputee subjects completed experiments without a physical prosthesis or load on the residual limb. It remains unclear if limb position affects real-time PR control in amputees when their residual limb is loaded with the weight of a prosthesis.

The aim of this experiment was to determine how limb position and external load affect real-time PR control in intact limb (ITL) and amputee (AMP) subjects and if these effects can be reduced with advanced training protocols. Subjects trained a controller 'statically', with the elbow bent at a 90° angle by their side, and 'dynamically', by moving the arm around their expected workspace. They then used the controllers to control a wrist and hand in a virtual reality (VR) environment to complete 3D Target Achievement Control (TAC) tests in four limb positions and with three external loads.

We hypothesized that limb position will negatively affect controller performance in ITL subjects but will not significantly affect AMP subjects when the residual limb is unloaded. In an

intact limb, forearm muscle activity changes to support the weight of the wrist and hand in different limb positions. Furthermore, the muscles that control wrist rotation change depending on elbow angle. As the elbow extends, the supinator receives less supination assistance from the biceps brachii, thus altering muscle activation patterns (Kendall & McCreary, 1983). Finally, there are several biarticular forearm muscles (brachioradialis, pronator teres, flexor carpi radialis, extensor carpi radialis longus, flexor carpi ulnaris, and palmaris longus) that assist with elbow flexion (Kendall & McCreary, 1983). In below-elbow amputees, forearm muscles are no longer attached to the wrist joint. Therefore, changes in muscle activity due to limb position would only be caused by the elbow-dependent supinator and the biarticular forearm muscles supporting the weight of the residual limb. As the residual limb is loaded, the increasing moment across the elbow may elicit further changes in these biarticular muscles. Therefore, we also hypothesized that the effects of limb position will be exacerbated in all subjects as the load increases. Furthermore, we hypothesized that increasing the load would decrease controller performance, as it has been shown to affect PR sensitivity (Cipriani et al., 2012). Finally, based on previous offline and real-time studies, we hypothesized that the dynamic training method will reduce the effects of limb position and load and improve controller performance (D. et al., 2017; Anders Fougner et al., 2011; Ashkan Radmand et al., 2014; Woodward & Hargrove, 2019).

2.3 Methods

This experiment was approved by the Northwestern University Institutional Review Board and all subjects gave written informed consent. We evaluated the effects of limb position, external load, and training method on PR-based myoelectric control in fourteen subjects with intact limbs

(ages 20-29, eight male, six female) and six subjects with major upper limb amputations (Table 2-1).

Table 2-1. Amputee Subject Demographics

Subject	Age	Gender	Time since amputation	Level of amputation	Home device control scheme	PR experience
AMP1	73	M	32 years	Transradial	Body-powered	Moderate
AMP2	33	M	5 years	Wrist Disarticulation	Direct control	Moderate
AMP3	65	M	6 years	Transradial	Body-powered	Moderate
AMP4	56	M	40 years	Transradial	Direct control	Moderate
AMP5	48	M	11 months	Transradial	Body-powered	None
AMP6	19	M	10 months	Transradial	Direct control	None

2.3.1 Experimental Setup

Intact Limb Subjects

Electromyographic (EMG) signals were recorded from ITL subjects using six pairs of equally spaced stainless-steel electrodes embedded in an armband. Each subject wore the armband around their right forearm, with the reference electrode just distal to the olecranon. An HTC Vive tracker was attached to the armband to track arm movements.

To encourage isometric contractions, subjects wore a forearm orthosis that restricted hand opening/closing and wrist flexion/extension. Orthoses that limit wrist rotation also obstruct free movement of the upper arm, which is required for the testing protocol, so we chose not to restrict wrist rotation. The external loads were secured to the palmar side of the orthosis using Velcro (Fig. 2-1A).

Amputee Subjects

EMG signals were recorded from AMP subjects using six pairs of wet electrodes equally spaced around the residual limb, avoiding regions above the radius and ulnar. The reference electrode was placed on the lateral epicondyle. A silicone liner was worn over the electrodes to prevent electrode shift or liftoff. Next, a lightweight frame was secured to the residual limb with a cuff, straps, and strong magnets. An HTC Vive tracker was attached to the cuff to track arm movements.

The external loads were placed into a box and screwed into the end of the frame. The frame was lengthened with modular extensions to position the load where a prosthetic hand would be. This was based on the distance from the lateral epicondyle to the middle of the hand on the intact side (Fig. 2-1B).

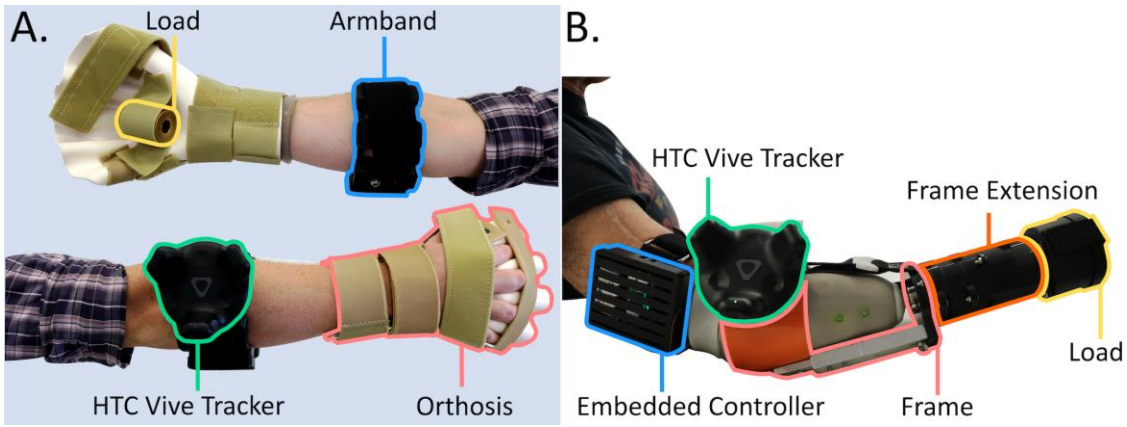

Figure 2-1. A: Palmar (top) and dorsal (bottom) view of the intact limb subject setup. The armband housed electrodes and an embedded controller that acquired and classified EMG signals. The orthosis was used to restrict wrist flexion/extension and hand opening/closing. Loads were attached to the orthosis. B: Dorsal view of the amputee subject setup. Wet electrodes were secured under a liner and signals were classified by an embedded controller. We used a frame to hold the tracker and loads, which were positioned where a prosthetic hand would be using frame extensions. The embedded controller was strapped to the upper arm and was not connected to the load-bearing frame. Both: The HTC Vive tracker tracked limb position.

2.3.2 Signal Acquisition and Processing

EMG signals were sampled at 1 kHz (Texas Instruments ADS1299), amplified with a hardware gain of 2, and a software gain of 1000, and band-pass filtered between 70-300 Hz. Data were then processed using an embedded System on Module (Logic PD SOMDM3730). We extracted four time-domain features (mean absolute value, waveform length, zero crossings, and slope sign changes) and six autoregressive coefficients from each channel using 200 ms windows in 25 ms increments. These features were used to train a linear discriminant analysis (LDA) classifier. We computed output velocity using previously described proportional control and velocity ramp algorithms (E. J. Scheme et al., 2014; Simon et al., 2011). After each processing window, the output class and velocity were wirelessly transmitted to a desktop computer and used to control an arm in a Unity-based virtual reality environment. The virtual arm was projected onto each subject's forearm using the location and orientation of the HTC Vive tracker.

2.3.3 Experimental Protocol

Each subject completed two sessions corresponding to two training strategies: static and dynamic (see descriptions of the training methods in the Section 2.3.4). The order of the sessions was pseudo-randomized such that half of the subjects used the static training method in the first session and the other half used the dynamic training method in the first session. In each session, subjects used the trained controller to complete VR-based TAC tests in four limb positions (Fig. 2-2) and with three loads (Table 2-2). To minimize the effects of learning, subjects were given at least two practice tests before actual data were collected. We chose limb positions that could be encountered in activities of daily living. Positions were constrained to the sagittal plane to ensure

that the loads would be supported by elbow and shoulder flexion only and to limit experimental time.

There were two physical interpretations of the external loads, summarized in Table 2-2. First, they corresponded to the average weights of different prosthesis configurations (Bajaj et al., 2015). Secondly, the 400g load represented the lightest possible powered wrist and hand prosthesis, while the 600g load represented the prosthesis grasping a small object.

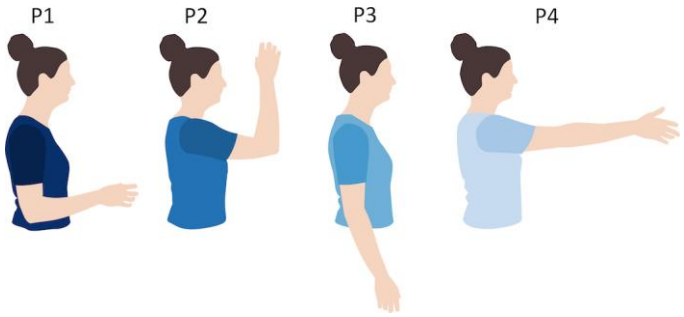

Figure 2-2. Limb positions used during testing protocol

Table 2-2. External Loads and Their Real-World Representations

Load (g)	Prosthesis Configuration		Prosthesis-Object Combination
	Wrist	Hand	
0	None	None	None
400	Passive	Powered	Prosthesis
600	Powered	Powered	Prosthesis + 200g

2.3.4 Training Data Collection

All ITL subjects and two AMP subjects (AMP1, AMP2) trained a 3DOF controller (hand open/close, wrist pronation/supination, wrist flexion/extension). The other four AMP subjects trained a 2DOF controller (hand open/close, wrist pronation/supination) because they were either naive to PR control (AMP5, AMP6) or unable to reliably perform wrist flexion and extension (AMP3, AMP4).

For each training method, we collected five sets of data per movement class. In each set, subjects were instructed to hold the movement for 2.5 seconds, resulting in 100 training patterns. AMP subjects used bilateral mirroring while training so that we could verify their movements. The training methods were as follows:

Static – Subjects kept their arm by their side with the elbow at a 90˚ angle (Fig. 2-2, P1) while holding the movement. There was no load attached. This is a common training method in experimental settings.

Dynamic – Subjects were instructed to move their arm freely around the workspace while holding each movement. They were encouraged to cycle through each test position (P1, P2, P4, P3 and reverse) by flexing and extending the elbow and shoulder. The 400g load, representing a lightweight prosthesis, was attached throughout training. This is an adaptation of a method suggested in previous studies (Ashkan Radmand et al., 2014; Woodward & Hargrove, 2019; D. Yang et al., 2017).

2.3.5 Real-Time Testing Protocol

In each session, subjects completed a set of 3D VR TAC tests that were based on previous work (Woodward & Hargrove, 2019). Each test comprised one trial for each movement class (except ‘no movement’) in each limb position. The test was repeated for each external load.

At the start of each trial, a target hand posture appeared in one of the four limb positions. Once the subject moved their arm within 10 cm and 30˚ of the correct limb position, the trial timer and myoelectric controller were activated. They then had 20 seconds to use the controller to move their virtual hand into the correct posture and remain there for 2 consecutive seconds. Each target was

52.5˚ away from the starting position and had a width of 16˚, corresponding to an index of difficulty of 2.10 bits (Fitts, 1954). Each target could be reached by using only one movement (hand open, wrist flexion, etc.). However, all trained DOF's were controllable. If a movement error was made, the subject had to use other DOF's to move back towards the target.

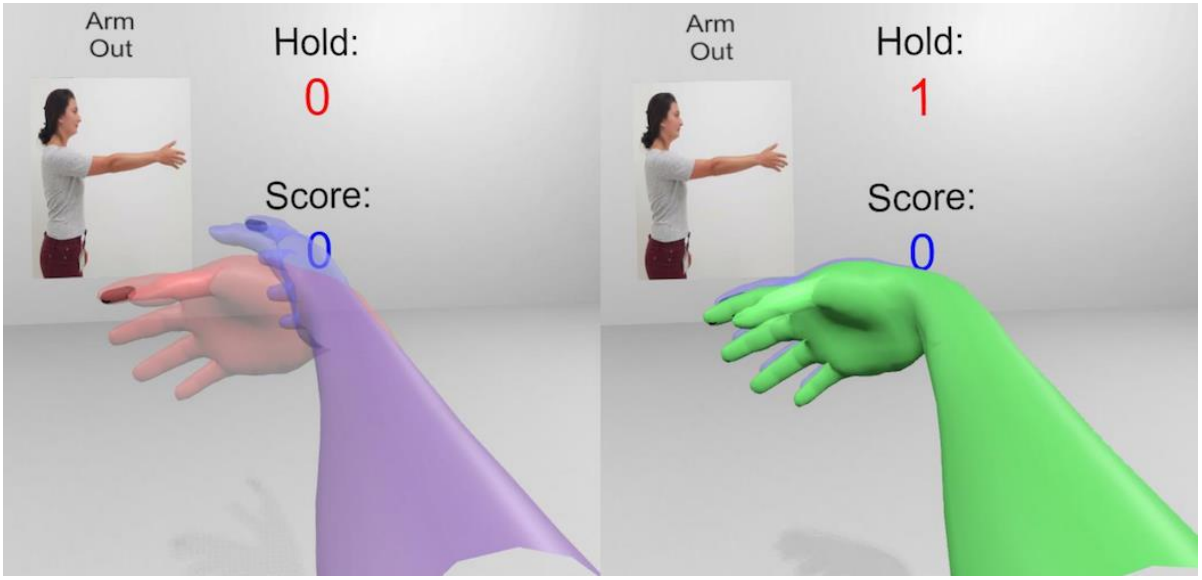

Figure 2-3. Example of a TAC test trial from the subject's view in the VR headset. In this trial, the user had to straighten their arm out (P4) and move the blue hand into the red target hand using wrist flexion (left). When the blue and red hands matched, the arm turned green (right).

2.3.6 Performance Evaluation

We quantified controller performance using the following outcome measures:

Completion rate (CR) – The percentage of trials that were successful within the allotted time limit. The trial was considered successful if they were able to remain in the target for 2 consecutive seconds. A completion rate of 100% meant that the subject was able to reach all targets.

Movement efficacy (ME) – The percentage of movements in each trial that were made towards the target, scaled by the proportional control output as shown in the equations below. At each

timepoint (n), given the current position of the virtual hand and the position of the target, we determined the class or classes (c_n^{target}) that needed to be activated to move the hand closer to the target. We summed the proportional control outputs (p_n^{out}) if their corresponding class output (c_n^{out}) matched c_n^{target}. We then divided this value by the sum of all proportional control outputs when the hand is not at rest ($c_n^{out} = 0$) or already in the target. This metric is an adaptation of path efficiency that considers the subject's ability to correct previous erroneous movements (Simon et al., 2011). A movement efficacy of 100% meant that the subject was able to move straight to and stay in the target without activating any other DOFs.

$$p_n^{direct} = \begin{cases} p_n^{out}, & \text{if } c_n^{out} \in c_n^{target} \\ 0, & \text{if } c_n^{out} = 0 \\ 0, & \text{else} \end{cases}$$

$$p_n^{all} = \begin{cases} 0, & \text{if } c_n^{out} = 0 \\ 0, & \text{if in target} \\ p_n^{out}, & \text{else} \end{cases}$$

$$ME = \sum_{n=1}^{N} \frac{p_n^{direct}}{p_n^{all}} \times 100\%$$

Stopping efficacy (SE) – The percentage of time that the virtual hand was at rest when it was in the target. A stopping efficacy of 100% meant that the subject was able to stop as soon as they reached the target.

Completion time (CT) – The time taken to complete each trial. For failed trials, the completion time was set to the time limit of 20 seconds. A lower completion time meant that the subject was able to reach the target faster.

2.3.7 Statistical Analyses

ITL and AMP data were analyzed separately. First, we determined how limb position and load affected the outcomes of each training method. We fit general linear models using a deviation effect coding scheme with limb position, load and their interaction as fixed factors and subject as a random factor. For AMP subjects, we added fixed factors for the number of trained DOFs and its interactions with limb position and load. This was used to determine if the reduced number of DOFs for AMP3, 4, 5, and 6 significantly affected their results. We used multiway analysis of variance (ANOVA) to quantify the statistical significance of each factor. Next, post-hoc pairwise comparisons were performed using Tukey-Kramer tests to quantify performance changes between specific positions and loads. Finally, to determine whether training method significantly affected our results, we pooled data from both sessions and fit a model using limb position, load, training method, and their interactions as fixed factors and subject as a random factor. Again, we used multiway ANOVA to quantify the significance of each factor. All p-values were adjusted using the Holm method to account for the number of outcome measures.

2.4 Results

2.4.1 Intact Limb Subjects

Mixed model results for fixed effects are presented in Table 2-3, in which shaded cells indicate statistical significance. For both training methods, the interaction terms between limb position and load were not statistically significant for any outcome measure, suggesting that the effects were independent. Thus, we analyzed the limb position effect pooled across all loads and vice versa (Fig. 2-4).

Static Training

Limb position significantly affected all outcome measures ($p < 0.001$). Depicted in Fig. 2-4, post-hoc results showed significant differences between almost all positions. As expected, subjects had significantly better control in the trained position (P1) than in all untrained positions (mean CR $\pm$ standard error: 77.38 $\pm$ 4.41%). P3 yielded the next best results (50.40 $\pm$ 6.26%), while P2 (37.30 $\pm$ 7.31%) and P4 (39.29 $\pm$ 6.63%) produced similarly poor results. These trends were consistent across all outcome measures.

Increasing the external load significantly decreased performance ($p < 0.001$). As shown in the pairwise comparison results in Fig. 2-4, completion rate, completion time, and movement efficacy were unaffected by the 400g load, but worsened with the 600g load. Stopping efficacy was the most sensitive to load, showing a significant decrease from 0g to 400g.

Dynamic Training

Based on Table 2-3, limb position effects were still statistically significant for most metrics ($p \leq 0.014$). However, dynamic training was able to significantly reduce these effects ($p < 0.001$). Limb position significantly affected completion rate, movement efficacy, and completion time, but not stopping efficacy ($p = 0.361$). Post-hoc results showed fewer statistically significant comparisons than the static training method. In particular, subjects performed equally between P1 and P3 (94.44 $\pm$ 2.40%, 93.25 $\pm$ 1.95%) and between P2 and P4 (86.90 $\pm$ 2.77%, 89.29 $\pm$ 2.88%).

Dynamic training did not significantly reduce the effects of load on any outcome measure ($p \geq 0.197$). Increasing the load still significantly affected all metrics negatively ($p \leq 0.003$).

Specifically, subjects generally performed significantly worse with the 600g load, but similarly with the 0g and 400g load.

Overall, the dynamic training method significantly improved control when compared to the static training method ($p < 0.001$). Subjects completed $90.97 \pm 2.10\%$ of all trials with the dynamic training method, in contrast to $51.09 \pm 4.93\%$ with the static training method. Notably, the outcomes at each position from dynamic training were better than outcomes at P1 from static training (the idealized case where training and testing were completed in the same position). Paired t-tests showed that these differences were all statistically significant for stopping efficacy and completion time ($p \leq 0.042$). Dynamic training also significantly improved completion rates ($p \leq 0.021$) at P1, P3, and P4 and movement efficacy ($p \leq 0.032$) at P1 and P3 compared to P1 from static training.

Table 2-3. Mixed Model Results for Intact Limb Subjects

		F-Value				*p*-Value			
		CR	ME	SE	CT	CR	ME	SE	CT
Static Training	Limb Position	44.85	57.67	51.99	46.30	<0.001	<0.001	<0.001	<0.001
	Load	13.45	12.01	9.13	13.23	<0.001	<0.001	<0.001	<0.001
	Limb Position*Load	0.80	0.50	0.40	0.74	1.000	1.000	1.000	1.000
Dynamic Training	Limb Position	4.07	8.60	1.07	6.87	0.014	<0.001	0.361	<0.001
	Load	8.35	11.83	5.98	13.28	<0.001	<0.001	0.003	<0.001
	Limb Position*Load	1.13	1.70	1.42	1.25	0.613	0.469	0.613	0.613
Overall	Training Method	556.67	654.58	1065.37	664.46	<0.001	<0.001	<0.001	<0.001
	Training Method*Limb Position	21.19	21.52	16.40	15.48	<0.001	<0.001	<0.001	<0.001
	Training Method*Load	3.02	0.68	0.17	0.92	0.197	1.000	1.000	1.000

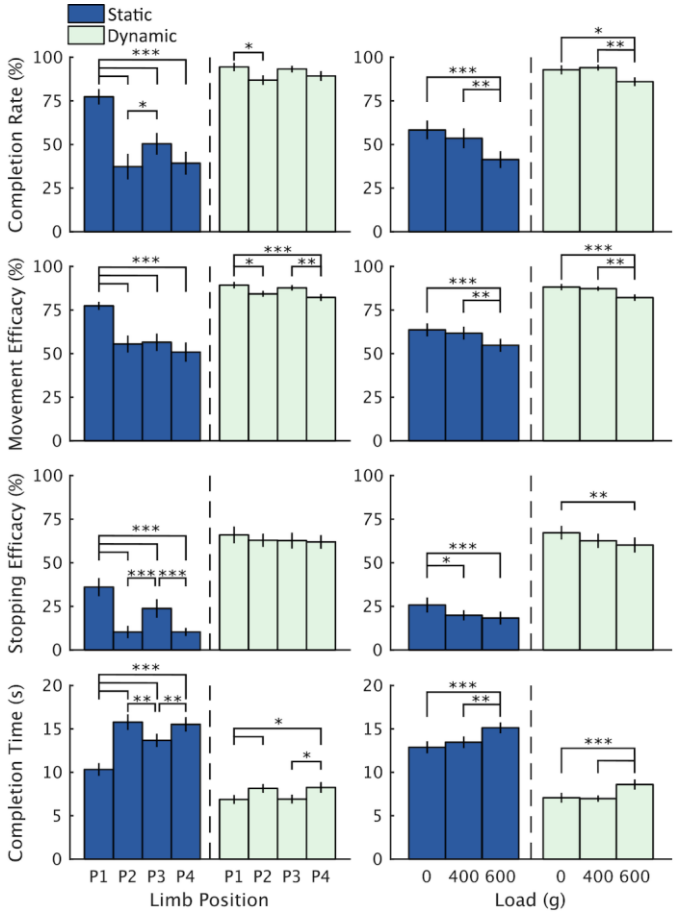

Figure 2-4. Intact limb subject results. Limb position (left) and load (right) significantly and independently affected controller performance regardless of training method. Dynamic training significantly mitigated limb position effects (and eliminated these effects in stopping efficacy), but did not reduce load effects. Overall, dynamic training significantly improved controller performance. Error bars represent standard errors.

2.4.2 Amputee Subjects

Mixed model results for fixed effects are presented in Table 2-4, in which shaded cells indicate statistical significance. Similar to the ITL results, all interaction terms between limb position and load were not statistically significant, indicating that the effects were independent. Thus, we analyzed the limb position effect pooled across all loads and vice versa (Fig. 2-5).

Static Training

Limb position significantly affected completion rate ($p = 0.004$) for all AMP subjects. Even though training only occurred in P1, subjects had similar control performances in P1, P2, and P3 ($79.62 \pm 8.42\%$, $76.85 \pm 8.07\%$, $65.28 \pm 12.08\%$). However, performance decreased in P4 ($58.33 \pm 14.13\%$). Limb position did not significantly affect movement efficacy ($p = 0.064$).

As shown in Table 2-4, the effects of limb position on stopping efficacy and completion time depended on the number of trained DOFs ($p \leq 0.007$). Hence, we evaluated these metrics separately for the two groups of AMP subjects. For 3DOF subjects, limb position affected the stopping efficacy and completion time at P4 ($p \leq 0.009$). For 2DOF subjects, limb position only affected stopping efficacy ($p = 0.020$).

Based on Fig. 2-5, increasing external load tended to decrease control performance. These changes were only statistically significant for movement efficacy ($p = 0.029$).

Dynamic Training

Limb position only significantly affected completion time ($p = 0.034$). Specifically, as shown in Fig. 2-5, subjects required more time to complete tasks in P4 compared to tasks in P1. Although dynamic training was able to eliminate position effects in all but one outcome measure, these changes were not statistically significant according to the interaction terms in Table 2-4 ($p \geq 0.311$).

Similar to ITL results, dynamic training did not reduce the effects of load ($p = 1.000$). External load significantly affected stopping efficacy ($p = 0.007$) but did not affect any other metric.

According to the interaction terms in Table 2-4, the impact of the training method was different for 3DOF and 2DOF subjects ($p \leq 0.003$). While dynamic training significantly improved overall control according to all metrics for 2DOF subjects ($p < 0.001$), it did not significantly affect control for 3DOF subjects ($p \geq 0.512$). However, 3DOF subjects had good control even with the static training method; they completed 86.11 $\pm$ 5.56% of all trials with the static training method and 90.28 $\pm$ 6.94% with the dynamic training method. In contrast, 2DOF subjects completed 61.98 $\pm$ 12.19% with the static training method and 83.33 $\pm$ 8.02% with the dynamic training method.

Again, the outcome measures resulting from dynamic training from all positions and loads were better or no different than the outcome measures from static training at P1 (the idealized case where training and testing were completed in the same limb position). However, unlike ITL subjects, these differences were not statistically significant.

Table 2-4. Mixed Model Results for Amputee Subjects

		F-Value				*p*-Value			
		CR	ME	SE	CT	CR	ME	SE	CT
Static Training	Limb Position	5.65	2.98			0.004	0.064		
	3DOF			5.13	10.42			0.009	<0.001
	2DOF			4.20	2.11			0.020	0.101
	Load	3.01	5.01	2.99	1.71	0.150	0.029	0.150	0.182
	Limb Position*Load	1.79	1.02	1.41	0.90	0.401	0.825	0.636	0.825
	Limb Position*DOF	1.89	3.08	4.82	5.10	0.130	0.056	0.008	0.007
	Load*DOF	0.59	0.11	2.76	0.03	1.000	1.000	0.259	1.000
Dynamic Training	Limb Position	1.54	2.16	1.41	3.97	0.407	0.113	0.407	0.034
	Load	0.54	1.99	6.45	2.20	0.586	0.338	0.007	0.338
	Limb Position*Load	0.83	1.91	0.81	2.05	1.000	0.234	1.000	0.234
	Limb Position*DOF	0.95	3.65	0.86	0.57	1.000	0.509	1.000	1.000
	Load*DOF	0.07		0.01	1.43	1.000	0.109	1.000	0.721
Overall	Training Method								
	3DOF	1.29	1.38	0.99	2.33	0.722	0.722	0.722	0.512
	2DOF	27.54	64.60	67.40	39.71	<0.001	<0.001	<0.001	<0.001
	Training Method*Limb Position	1.21	0.16	2.29	0.87	0.918	0.925	0.331	0.918
	Training Method*Load	0.82	0.44	0.22	0.03	1.000	1.000	1.000	1.000
	Training Method*DOF	9.16	22.19	20.51	10.36	0.003	<0.001	<0.001	0.003

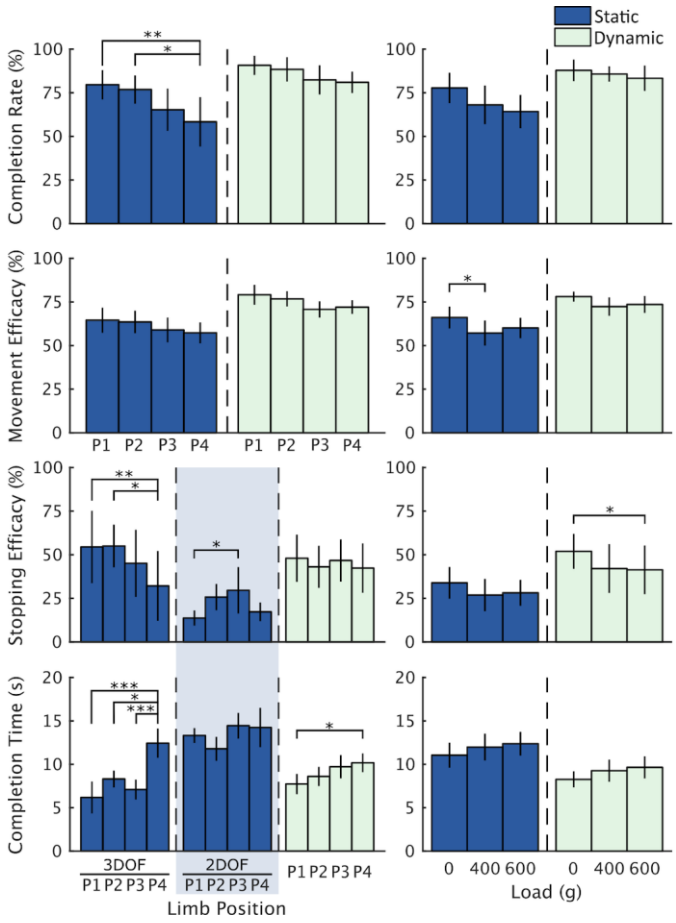

Figure 2-5. Amputee subjects results. When the controller was trained statically, limb position (left) significantly affected completion rate for all amputee subjects. Limb position affected completion time and stopping efficacy for subjects who used a 3DOF controller, and affected stopping efficacy for subjects who used a 2DOF controller. There were fewer differences between positions than in intact limb subjects. When dynamic training was used, limb position only affected the completion time of P4. Although dynamic training reduced limb position effects, this change was not statistically significant. Increasing the load (right) significantly decreased movement efficacy and stopping efficacy. Overall, training dynamically significantly improved controller performance.

2.5 Discussion

Previous work showed that limb position impacted offline PR classification rates in ITL (Betthauser et al., 2018; A Fougner et al., 2011; Khushaba et al., 2014; Liu et al., 2014; E. Scheme et al., 2011) and AMP populations (Y. Geng, Zhou, et al., 2012; L. Chen, Y. Geng, 2011) as well

as real-time outcome measures in ITL subjects (Betthauser et al., 2018; Khushaba et al., 2014; Woodward & Hargrove, 2019). However, real-time limb position effects were minor in AMP subjects when the residual limb was unloaded (Woodward & Hargrove, 2019). These effects were not well understood and it was unknown how they would change if the residual limb was loaded. In this study, we addressed this gap by evaluating the effects of limb position and load on real-time PR control in ITL and AMP subjects. We also evaluated if a dynamic training method, adapted from (Ashkan Radmand et al., 2014; Woodward & Hargrove, 2019; D. Yang et al., 2017), would be able to mitigate these effects.

We found that limb position and load independently affected control for all subjects. Furthermore, the dynamic training method significantly improved controller performance compared to the static training method. This was expected because the dynamic training method encompassed more of the test workspace and increased training data variability. There were notable differences between each subject population.

Intact Limb Subjects

The results from the ITL subjects were consistent with previous findings showing that controllers were significantly affected by limb position when trained statically. All positions had significantly worse results compared to the position in which the controller was trained (P1). Subjects had the most difficulty in positions where elbow stabilization required more effort (P2 and P4). Many often commented that the controller “drifted” towards wrist supination. This was likely caused by increased activity in the brachioradialis, which assists with elbow flexion and also controls forearm pronation/supination. Although several other forearm muscles help with elbow flexion, the brachioradialis provides the most assistance. It is also more superficial and just distal to the elbow,

close to where the electrodes were located. In extended elbow positions, this drift may have also been caused by changes in supinator activity. The drift was less noticeable in the dynamic sessions, suggesting that moving the arm throughout the workspace during training accounted for the changes in brachioradialis and supinator activity due to elbow flexion.

Even though dynamic training significantly improved results, limb position still impacted control, albeit in a less substantial way. Specifically, the completion rate range between limb positions in the static session was 40.08%, while the range from the dynamic session was 7.54%. Therefore, although it was not eliminated, the limb position effect was reduced significantly when the controllers were trained dynamically.

Notably, stopping efficacy was not affected by limb position after dynamic training. Previous work found that external loads increased baseline muscle activity and affected the classification of 'no movement' (Cipriani et al., 2012). By training with a load and with arm movements, we increased the threshold of the 'no movement' class and reduced the likelihood of triggering unwanted movements in all limb positions.

Amputee Subjects

The results from the AMP subjects did not reiterate all the trends seen in ITL subjects or previous studies with offline analyses. P4 was the only limb position that significantly diminished controller performance, as opposed to P2, P3, and P4 in ITL subjects. AMP subjects were able to complete 58.33% of the trials in P4 in the static session. Although this was the worst completion rate for AMP subjects, it was still better than ITL subject performance in P2, P3, and P4. Similar differences between subject populations were observed during offline prediction of kinematics (Ning Jiang et al., 2013). These findings may be explained by the anatomical differences described

in the introduction and the cited work. Unlike in ITL subjects, forearm muscles in AMP subjects were only affected by elbow joint moments. When the elbow was extended straight out (P4), the moment across it was maximized. This, in turn, influenced activation patterns that involved the brachioradialis and the supinator, thus reducing controllability. These changes may have been more extreme in ITL subjects because they also had to support wrist moments created by the weight of an intact hand.

Dynamic training eliminated all the effects of P4 for all metrics except completion time. Although we instructed subjects to flex and extend their elbow and shoulder during the dynamic training method, we did not enforce strict movements. Consequently, subjects had a tendency to stay within more comfortable positions during training and avoided straightening their arm out (P4). In the future, a more controlled training protocol that requires subjects to keep their arm straight out could be imposed. Alternatively, users could be instructed on how to compensate for residual limb loading by calibrating in the regions that are important for prosthesis usage (e.g. prosthesis guided recalibration (Simon et al., 2012). We believe that these training methods would be sufficient to mitigate the effects of limb position.

Although there was a performance decrease as load increased, these changes were, for the most part, not statistically significant. Many of these subjects were consistent prosthesis-users and were therefore accustomed to supporting loads on their residual limb. However, given that we tested only three loads, more research is required to determine precisely how limb loading can impact performance as weights increase.

2.5.1 Limitations and Future Work

One of the major limitations of this study was the influence of muscle fatigue. Although we allowed subjects to take frequent breaks, many remarked that the weight began to feel heavier over time. As fatigue set in, subjects had difficulty keeping their arm in the correct limb position, thus affecting test metrics such as completion rate and time. Therefore, changes in these metrics may not solely have been caused by changes in myoelectric control performance. Fatigue may have also affected the repeatability of control signals, which would have been apparent in the movement and stopping efficacy metrics.

Although we found statistical significance in some results, we do not know how this correlates to clinical or functional significance. The TAC test requires a level of precision that may be more than what is necessary for prosthesis usage in many tasks. For example, the virtual hand needs to stop halfway between opening and closing, which may be necessary when grasping a delicate object, but not required for many other objects. Hence, statistically significant effects on TAC test metrics may not be clinically significant. Future work should extend this work to include functional tests that are more clinically applicable.

Furthermore, we only had six AMP subjects. There was a wide range of skill levels; some subjects had participated in many previous experiments while others had never used PR before. With the small sample size and large variance, AMP subject data had lower statistical power than ITL subject data. This may explain the statistical differences between AMP and ITL subjects. However, as previously mentioned, statistical significance does not imply clinical significance.

Four AMP subjects used a 2DOF controller while all other subjects used a 3DOF controller. We found that this significantly impacted the limb position effect on stopping efficacy and completion

time. With more decision boundaries, the 3DOF controller may have been more susceptible to misclassifications. This may have been exacerbated by limb position and load, resulting in bigger effects. Subjects who used the 2DOF controller also had fewer trials and may have experienced less fatigue. Nonetheless, the limb position effect in both AMP groups were still different from that in ITL groups.

Our conclusions were also limited by the number of limb positions and loads that were tested. By constraining the limb positions to the sagittal plane, we isolated the effects of shoulder and elbow flexion/extension on controller performance. However, control may also be affected by limb positions that require abduction/adduction, humeral rotation, and other DOF's. Furthermore, we chose to limit our heaviest load to 600g to prevent fatigue. In reality, individuals may use heavier prostheses or carry larger objects. Although we did not find any significant interactions between limb position and load, this result may not hold with different limb positions and heavier loads.

Since we only used PR-based controllers, we may not be able to generalize our conclusions to other myoelectric controllers. However, our AMP subject results were consistent with the results from studies with regression-based controllers (Janne M. Hahne et al., 2018; Hwang et al., 2017). Specifically, both studies found that real-time control in AMP subjects was mainly affected by the 'arm out' position (P4). Our load effects were larger than the effects on a musculoskeletal model-based controller (Pan et al., 2018). However, they did not include wrist rotation, which was the most affected DOF in our study.

Finally, our electrode and load-bearing setup was not representative of a typical socket with embedded dry electrodes. With a clinically prescribed socket, different limb positions and loads may cause electrode shift or liftoff. Even though it may not have accurately reflected real-world

conditions, we chose to use our setup so that we could isolate the physiological effects of limb position and load. Although ITL subjects used dry electrodes, we ensured that the armbands were worn tightly and that the electrodes did not shift throughout the experiment.

2.6 Conclusion

In this study, we evaluated the effects of limb position, load, and training method on real-time pattern recognition control in intact limb and amputee subjects. We showed that limb position and load significantly and independently affect controller performance when static training is used. However, these effects are different in AMP subjects and ITL subjects and may not be as pronounced as previous offline results have shown. This highlights the importance of testing controllers in real-time with their intended end user. We also showed that increasing external loads diminishes controllability, although these changes may not be statistically significant in AMP subjects. Finally, dynamic training reduces the effects of limb position but not load and substantially improves overall controller performance. In the future, the dynamic training method should be adapted to require more training with the arm straight out, since this was found to be the most problematic position.

3 Data Augmentation and Deep Learning Techniques to Improve Interface Noise Tolerance of Myoelectric Pattern Recognition Controllers

3.1 Abstract

Background: Clinically available myoelectric PR controllers deteriorate under conditions that generate interface noise, such as electrode liftoff or wire failure. Previous solutions relied on additional processing steps like signal denoising and controller adaptation to mitigate these negative effects. However, there are no clinically practical controllers that are inherently robust to interface noise. This paper investigated the use of data-driven methods to build clinically practical, noise tolerant PR control strategies for transradial prostheses.

Methods: We developed a data augmentation protocol to increase training data variability and two deep neural networks (MLP and CNN) to compress six-channel EMG inputs to a latent space that is linearly separable and robust to noise. We evaluated seven gesture classification strategies – four were trained on standard training data sets (LDA, LDA-, MLP-LDA, CNN-LDA) and three were trained on augmented training data sets (LDA+, MLP-LDA+, CNN-LDA+). LDA, LDA+, and LDA- used LDA classifiers to classify time domain EMG features while MLP-LDA, MLP-LDA+, CNN-LDA, and CNN-LDA+ aligned the features to the MLP and CNN latent spaces before classifying them with LDA classifiers. Using data from fourteen intact limb subjects and seven below-elbow amputee subjects, we computed the classification accuracies on clean signals and signals that had interface noise in up to four channels.

Results: Compared to the baseline LDA method, LDA+ and MLP-LDA+ significantly increased classification accuracy on noisy data but decreased accuracy on clean data. On the other hand, LDA- and CNN-LDA+ improved noise tolerance while maintaining high performance with clean data.

Conclusion: The CNN-LDA+ classification strategy is robust to multichannel interface noise and can improve the reliability and usability of PR-based upper limb prostheses.

3.2 Background

Major upper limb amputations cause impairments that inhibit basic activities of daily living such as eating and dressing (Cordella et al., 2016). Myoelectric prostheses have the potential to restore some lost functionality, thereby helping users to regain their independence and improve their quality of life. Commercial devices employ two main types of control systems: amplitude-based direct control and pattern recognition (PR) control.

Direct control uses electromyographic (EMG) signals measured from agonist-antagonist muscle pairs to control prosthesis movements (Williams et al., 2004). Each muscle pair activates one degree-of-freedom (DOF); thus, the dexterity of a direct controlled prosthesis is limited by the number of distinct muscle sites that can be individually contracted. Although co-contractions can be used to toggle between DOFs, this renders the control method less intuitive. Furthermore, while direct control is reliable, it requires manual tuning to accommodate the unique muscle anatomies of each user.

PR control methods measure EMG signals from an array of electrodes and learn the patterns of muscle activity that correspond to intended movements (Englehart & Hudgins, 2003; Hudgins et al., 1993; Parker et al., 2006). Typically, PR controllers use a classifier to classify descriptive features extracted from windowed EMG signals. Use of linear discriminant analysis (LDA) classifiers is common as they are computationally simple to train and implement (Coapt LLC, n.d.; Ottobock, n.d.; Simon et al., 2019). PR control has been shown to improve functional outcomes for prosthesis users, provided the EMG interface is stable and the signals are repeatable (L. J. Hargrove et al., 2013).

However, regular usage of myoelectric prosthesis gives rise to various sources of signal disturbances that degrade classification accuracy (Chowdhury et al., 2013; De Luca et al., 2010; Kyranou et al., 2018; E. J. Scheme & Englehart, 2011; Simon et al., 2012; Zhang & Huang, 2015). Changes in residual limb volume, limb position, and socket loading can cause electrodes to intermittently lose contact with the skin. Signal abnormalities stemming from electrode or wire failures also occur with prolonged prosthesis use (Levi Hargrove et al., 2018; Miller et al., 2020; Simon et al., 2012). Interface noise in just one channel is often detrimental to the accuracy of PR control strategies, rendering the device unusable until it can be recalibrated (Simon et al., 2012; Zhang & Huang, 2015).

Previous studies have proposed several approaches to resolve the effects of interface noise. For example, signal processing algorithms can be used to denoise affected channels before classification (De Luca et al., 2010; Fraser et al., 2011; Maier et al., 2018; Ortolan et al., 2003; Phinyomark et al., 2009; Powar et al., 2018; Reaz et al., 2006; Rehbaum & Farina, 2015). The generalizability of these methods is limited, however, as most focus on filtering out periodic noise (eg. electrical noise) and do not address the effects of intermittent noise signals (Phinyomark et al., 2020). Another approach uses control strategies that adjust their classifier parameters to adapt to changes in EMG signals (Sensinger et al., 2009; Stachaczyk et al., 2020; Tommasi et al., 2013). Of note is a fast-retraining LDA method that detects and removes noisy channels before recalibrating its weights (Zhang & Huang, 2015). Though they can increase classification accuracy, these adaptive control methods require additional processing steps during classification.

One promising solution exploits the signal redundancies across EMG channels to build a classifier that is inherently robust to interface noise. Since surface electrodes measure diffuse muscle activity, an array of surface EMG signals contains overlapping neural information. These

redundant signals can therefore be compressed into a low-dimensional manifold that retains discriminative features and is less sensitive to input disturbances (Gazzoni et al., 2004; Ison & Artemiadis, 2014; N. Jiang et al., 2009; Teh & Hargrove, 2021). Using this concept, prior works have employed spatial filtering, linear factorization, and data fusion to build robust classifiers (Janne Mathias Hahne et al., 2012; López et al., 2009; Muceli et al., 2014). However, these examples used setups that are not yet clinically practical (eg. high-definition EMG arrays).

To train an accurate PR controller, it is essential that the training data typifies signals in real scenarios. For example, performing dynamic arm movements during training data collection instead of maintaining a static position significantly improves LDA classification performance across different limb positions (Anders Fougner et al., 2011; Teh & Hargrove, 2020). By increasing training data variability, classifiers are encouraged learn discriminative features that are consistent across sources of variance, thus preventing overfitting. However, it is a challenge to physically collect enough data to sufficiently represent realistic scenarios. To alleviate the burden of extensive training data collection, data augmentation can be used to artificially introduce variability in a systematic manner (Hu et al., 2019; Luo et al., 2019; Shorten & Khoshgoftaar, 2019; D. Yang et al., 2021).

As the variance of the training data increases and the data become more complex, simple linear classifiers may not be equipped to adequately model hidden structures. Deep learning models are known for their ability to learn complex nonlinear relationships within large datasets. Specifically, deep encoders can find invariant features across noisy high-dimensional inputs and compress them into robust low-dimensional latent subspaces (Lecun, 1987). A desirable characteristic of these manifolds learned from noisy data is that large disturbances in the input have minimal effects on their latent representations. Depending on the model's loss function, the

latent space can be optimized for a specific objective, such as reconstructing or classifying the input.

Convolutional neural networks (CNNs) are another useful deep learning tool. CNNs are commonly used for image processing applications where the relative locations of pixels are crucial to the underlying structure. While sequential neural networks use flattened vectors as inputs and process all input pixels the same way, CNNs allow multidimensional input matrices and use kernels to process pixels based on their locality. This emphasizes the local structure of the data and regularizes the model through sparsity of connections and shared parameters (Holden et al., 2015; Krizhevsky et al., 2012; Lecun & Bengio, 1995). Thus, CNNs are well-equipped to disentangle EMG signal redundancies, which are dependent on electrode locations (Ameri et al., 2019; Atzori et al., 2016; W. Geng et al., 2016; Luo et al., 2019; W. Yang et al., 2019).

The objective of this study was to develop a clinically feasible, noise-tolerant myoelectric PR controller that classifies hand and wrist movements. To that end, we explored the use of data augmentation and deep learning techniques to uncover a latent subspace in which movement classes are separable and robust to interface noise. We trained seven gesture classification strategies and evaluated their performances on normal EMG data and EMG data that contained up to four channels of interface noise. These strategies included LDA algorithms that classified time domain EMG features and LDA algorithms that classified latent EMG variables computed by an MLP and a CNN. We hypothesized that the CNN classification strategy trained with augmented data would be the most accurate non-adaptive classifier across all noise conditions because it preserves spatial dependencies and contains nonlinearities.

The paper is organized as follows. First, we describe the experimental protocol, classification schemes, data analysis, and statistical measures in the methods section. Next, we

present the offline performances of the gesture classification strategies, latent representation plots, and classifier computation time in the results section. Thereafter, we discuss the implications of and possible explanations for the results, as well as the limitations of the study. Finally, this paper finishes with summarized findings and concluding thoughts.

3.3 Methods

The following experiment was approved by the Northwestern Institutional Review Board. Fourteen individuals with intact limbs (ITL) and seven individuals with below-elbow amputations (AMP, Table 3-1) participated in this study after providing written informed consent. Due to partial data loss, results from one ITL participant and one AMP participant were excluded from the final analysis.

Table 3-1 Amputee Subject Demographics

Subject	Age	Gender	Time Since Amputation	Level Of Amputation	DOFs Controlled
AMP1	73	M	32 years	Transradial	3DOF
AMP2	33	M	5 years	Wrist disarticulation	3DOF
AMP3	65	M	6 years	Transradial	2DOF
AMP4	56	M	40 years	Transradial	2DOF
AMP5	48	M	11 months	Transradial	2DOF
AMP6	19	M	10 months	Transradial	2DOF

3.3.1 Experimental Setup

For ITL participants (Fig. 3-1a), six channels of EMG signals were collected using dry stainless-steel bipolar electrodes (Motion Control Inc.) that were embedded in an adjustable armband. The electrodes were equally spaced around the subject's right arm, with the reference electrode positioned just distal to the olecranon. An HTC Vive tracker was attached to the dorsal side of the armband and used to track the participant's limb position. Participants also wore an orthosis around the wrist and hand to promote isometric contractions that would more closely

resemble amputee contractions. Finally, a 400g weight was attached to the distal end of the orthosis to simulate the weight of a prosthesis.

Due to the unique size and anatomy of each residual limb, dry electrode setups that are not specifically customized for an amputee are prone to electrode liftoff. Hence, we used wet electrodes for AMP participants to prevent unwanted interface noise (Fig. 3-1b). Six channels of EMG signals were collected using adhesive Ag/AgCl bipolar electrodes (Bio-Medical Instruments) that were secured under a silicone liner. The electrodes were equally spaced around the subject's residual limb and the reference electrode was placed just distal to the olecranon. An adjustable lightweight frame was fastened around the residual limb and lengthened to match the subject's intact limb length. A 400g weight was attached to the distal end of the frame to simulate the weight of a prosthesis.

3.3.2 Data Collection Protocol

All data collection was conducted in an HTC Vive virtual reality environment (Fig. 3-1c). Each participant collected a training data set and a test data set during one experimental session. EMG signals were sampled at a rate of 1 kHz, band-pass filtered between 70-300 Hz, and segmented into 200 ms windows in 25 ms increments. In addition to a hardware gain of 2 and a software gain of 1000, there were channel-specific software gains that were customized for each subject. These channel gains were calculated by scaling the signals in each channel to span the output range of -5V to 5V. Although the channel gains were calculated using the training data set alone, they were applied to the training and test data sets.

To collect training data, the subject performed hand and wrist gestures while moving their arm around the workspace. This simple training protocol has been shown to achieve high real-time performance (Anders Fougner et al., 2011; Teh & Hargrove, 2020). All ITL subjects and two AMP

subjects completed seven gestures (rest, wrist flexion/extension, wrist pronation/supination, hand open/close), corresponding to a 3DOF controller. Based on clinician input and to minimize fatigue, the remaining four AMP participants completed five gestures (rest, wrist pronation/supination, hand open/close), corresponding to a 2DOF controller. Each gesture was held for 2.5 seconds and repeated five times, resulting in 12.5 seconds (500 overlapping windows) of clean training examples per gesture.

To collect test data, the subject performed the trained hand and wrist gestures in four limb positions (Fig. 3-3a). Each gesture was held for 2.5 seconds and repeated five times. Therefore, each participant had 50 seconds (2000 overlapping windows) of clean test data for each gesture.

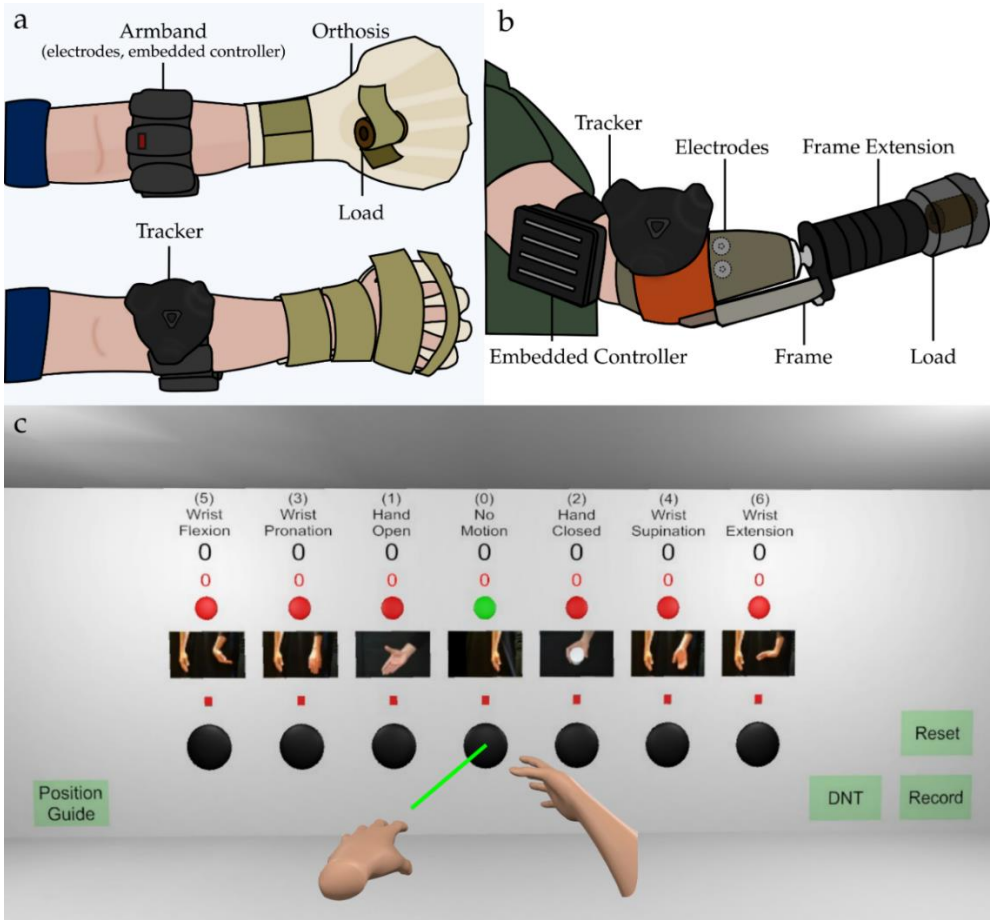

Figure 3-1. Experimental setups for (a) intact limb participants and (b) amputee participants. (c) Virtual reality environment used for data collection.

3.3.3 Offline Analyses

After EMG data collection, all further analyses were conducted offline on a Windows 10 laptop computer with 16GB RAM, an Intel Core i7-9850H CPU at 2.60GHz, and a 4GB NVIDIA Quadro T1000 GPU. These analyses included training data augmentation, training five gesture classification strategies, testing those strategies, and statistical evaluations.

3.3.4 Training Data Augmentation

We constructed an augmented training data set by systematically corrupting up to four EMG channels in copies of the original raw training signals (Fig. 3-2). The augmented data set can be divided into five subsets, each containing a different number of corrupted channels (0 to 4). The subset with zero corrupted channels was simply a copy of the original training data set. For each of the remaining subsets, we first created a reference subset composed of two copies of the original training data set. We partitioned the reference subset into ${}_6C_n$ segments, designating a distinct combination of channels for each segment. We then evenly distributed 12 types of synthetic noise into the designated channel combinations. These synthetic noise types included flatlining, in which the signal was completely attenuated to 0V, five levels of Gaussian noise centered at 0V (σ = 1, 2, 3, 4, 5V), five levels of 60 Hz noise (amplitude = 1, 2, 3, 4, 5V), and a randomized mixture of all noise types. In summary, the augmented training data set was nine times the size of the original training data set and contained 12 types of synthetic noise in all combinations of zeros to four channels.

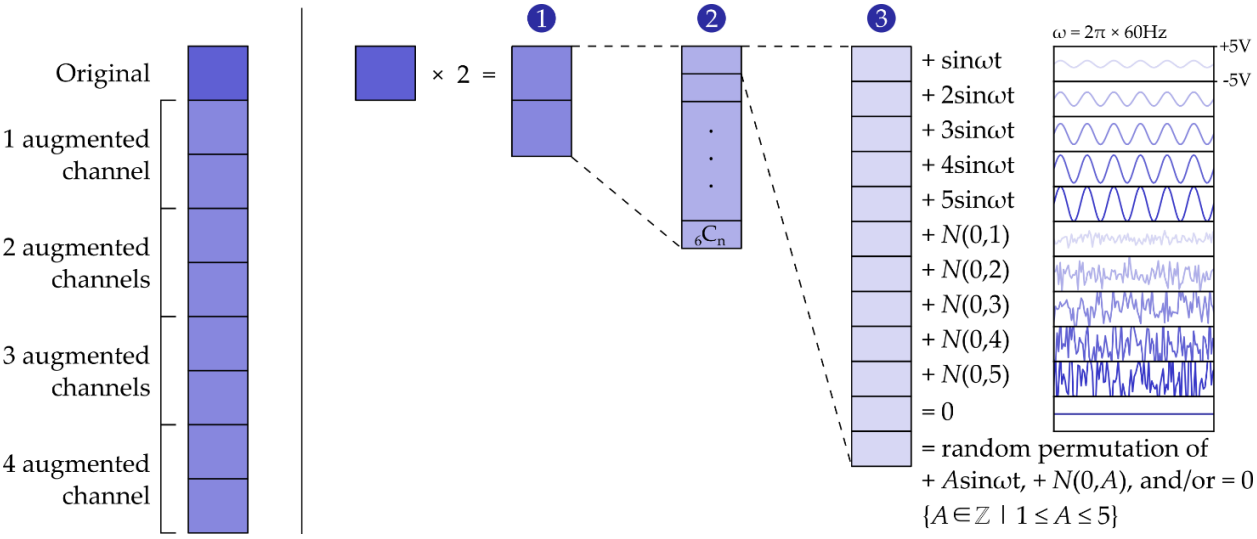

Figure 3-2. The augmented data set contained one copy of the original training data and eight copies of the training data that were synthetically corrupted in up to four channels. The steps to create an augmented data subset with n corrupted channels were as follows: (1) Two copies of the original training data were made. (2) The data were partitioned into 6Cn equal segments, with each segment containing a distinct combination of n corrupted channels. (3) These channels were augmented by adding 60Hz noise (amplitude = 1-5V), adding Gaussian noise (μ = 0, σ = 1-5V), attenuating to zero, or a random permutation of the aforementioned.

3.3.5 Gesture Classification Strategies

Seven gesture classification strategies were trained and evaluated in this study. Before we trained the algorithms, four time-domain features were extracted from the training data sets: mean absolute value, waveform length, zero crossings, and slope sign changes (Hudgins et al., 1993).

Traditional LDA Classifiers

Three control methods were based on the traditional LDA classifier algorithm.

1. Baseline LDA (LDA) – To act as the baseline model, an LDA classifier was trained with the original training data set. This algorithm is used in most clinically available PR systems and therefore demonstrates what current prosthesis users experience.

2. Augmented LDA (LDA+) – To investigate how data augmentation affects the reliability of a standard LDA algorithm, we trained an LDA classifier with the augmented training data set.
3. Adaptive LDA (LDA-) – To provide a benchmark for adaptive solutions, we implemented an existing fast-retraining LDA classifier that was shown to successfully reduce the effects of interface noise offline and in real-time (Zhang & Huang, 2015). This system used an automatic fault detector to detect channels that contained signal noise, then removed the noisy EMG channels and adjusted the LDA weights to match the new channel configuration. For our simplified implementation, we omitted the fault detector module and used the noise label ground truth to simulate a perfect fault detector, thus providing the best-case scenario for the adaptive method.

   First, we trained an LDA classifier with the original training data set and stored the class mean and covariance matrices. When classifying a noisy sample, we first removed the noisy channels from the input vector to form a reduced input vector. Then, we remove the rows and columns in the mean and covariance matrices corresponding to the noisy channels. We used these mean and covariance matrices to calculate new weights for the LDA classifier, which then classified the reduced input vectors. When there were no noisy channels, the LDA- classifier was identical to the baseline LDA classifier.

Neural Network-Aligned Classifiers

The remaining four classification strategies comprised two stages: a latent encoder network that aligned the EMG inputs to a low-dimensional manifold and an LDA classifier that classified these latent variables (Fig. 3-3d). Two strategies used a multilayer perceptron network (MLP-LDA, MLP-LDA+) while the other two used a convolutional neural network (CNN-LDA, CNN-

LDA+). We chose to classify the latent variables using LDA classifiers instead of the neural networks' linear classifier layer to maintain a consistent classifier across all classification strategies. This allowed us to compare the LDA components across strategies and visualize the class separability in the latent spaces. Both models were implemented using Keras 2.3.1 with the Tensorflow backend (Chollet & others, 2015).

We used five-fold cross-validation to tune model hyperparameters (Table 3-2). To avoid overlapping training and validation data, each fold corresponded to one gesture repetition. We based the initial model architectures and hyperparameters on similar work that used deep convolutional networks for prosthesis control. Then, we performed manual tuning with a coordinate descent method starting with the learning rate, followed by (Bengio, 2012; Hu et al., 2019). After the hyperparameters were determined, we trained the final models with the entire augmented training data set.

The networks were trained using the Adam optimization algorithm with a learning rate of 0.001 and mini-batch gradient descent with 30 training epochs and a batch size of 128 (Kingma & Ba, 2015). To accelerate training time, we used a minmax scaler to standardize the input features between [0,1] and applied batch normalization after each hidden layer (Ioffe & Szegedy, 2015). To encourage sparsity, L1 regularization (λ = 1e-5) was applied to all hidden layers (Ng, 2004).

Table 3-2. List of Hyperparameter Values for MLP and CNN

Hyperparameter	Value
Adam optimizer	$\beta_1 = 0.9$, $\beta_2 = 0.900$, ε = 1e-7
Learning rate	0.001
Mini batch size	128
Training epochs	30
L1 regularization	λ = 1e-5
Loss function	Categorical crossentropy

4. Multilayer perceptron-aligned LDA (MLP-LDA) – Using the original training data, we trained a fully connected five-layer neural network (Fig. 3-3b) to take in a 24 by 1 EMG feature vector and output a 4 by 1 latent feature vector z_{mlp} and a predicted gesture label $\hat{y}_{mlp}$.

   The first four hidden layers aligned the EMG inputs to the latent space. We applied ReLU activation functions after the first three layers and a linear activation function after the fourth layer (Nair & Hinton, 2010). Based on our cross-validation results, we found that classification accuracy improved as the dimensionality of the latent space increased but began to plateau after a dimensionality of 4. Thus, we set the latent dimension to 4.

   The last hidden layer in the MLP was a linear classifier that used a softmax activation function to classify the latent features. The network was trained to minimize the categorical cross entropy loss between the predicted class and the ground truth, thus optimizing linear separability between movement classes in the latent space.

   We trained an LDA classifier with the latent features of the augmented data set and used it in tandem with the MLP network to form the MLP-LDA classification strategy (Fig. 3-3d). During classification, the EMG input vectors were passed through the MLP encoder to compute the latent features z_{mlp}, which were then fed to the LDA classifier to obtain gesture predictions.

   In total, the MLP had 1267 trainable parameters.

5. Augmented MLP-LDA (MLP-LDA+) – We trained the MLP-LDA model (Fig. 3-3) with the augmented training data set to examine the effects of data augmentation on an MLP-based classification strategy.

6. Convolutional neural network-aligned LDA (CNN-LDA) – We trained a CNN (Fig. 3-3c) on the original training data set with the same objectives as the MLP: to output a 4 by 1 EMG latent feature vector z_{cnn} and a predicted gesture label $\hat{y}_{cnn}$.

   While the inputs for the previous classifiers were 24 by 1 feature vectors, the CNN input was a 6 by 4 feature matrix, corresponding to the 6 EMG channels and 4 time-domain features. This enabled the 2-dimensional convolutional layers to exploit the spatial relationships between EMG channels and learn more robust latent representations.

   The first five hidden layers served as the encoder, starting with two convolutional layers with ReLU activation functions. Then, we flattened the output of the convolutional layers before passing it through two sequential layers with a ReLU and a linear activation function, respectively. Thus, the latent encoder modules of the CNN and MLP each had three ReLU functions and one linear function. Like the MLP, the latent space had a dimensionality of 4.

   The last layer of the CNN classified the latent feature vector z_{cnn} using a softmax activation function. The CNN was trained to minimize the categorical cross entropy loss between the predicted class and the ground truth, once again to encourage linear separability between the class latent representations.

   Finally, we trained an LDA classifier with the augmented training data set after it was aligned by the CNN. The CNN-LDA classification strategy (Fig. 3-3d) used the CNN encoder to compute latent variables z_{cnn} which were then classified with the LDA classifier. In total, the CNN had 12999 trainable parameters.

7. Augmented CNN-LDA (CNN-LDA+) – We trained the CNN-LDA model (Fig. 3-3) with the augmented training data set to examine the effects of data augmentation on a CNN-based classification strategy.

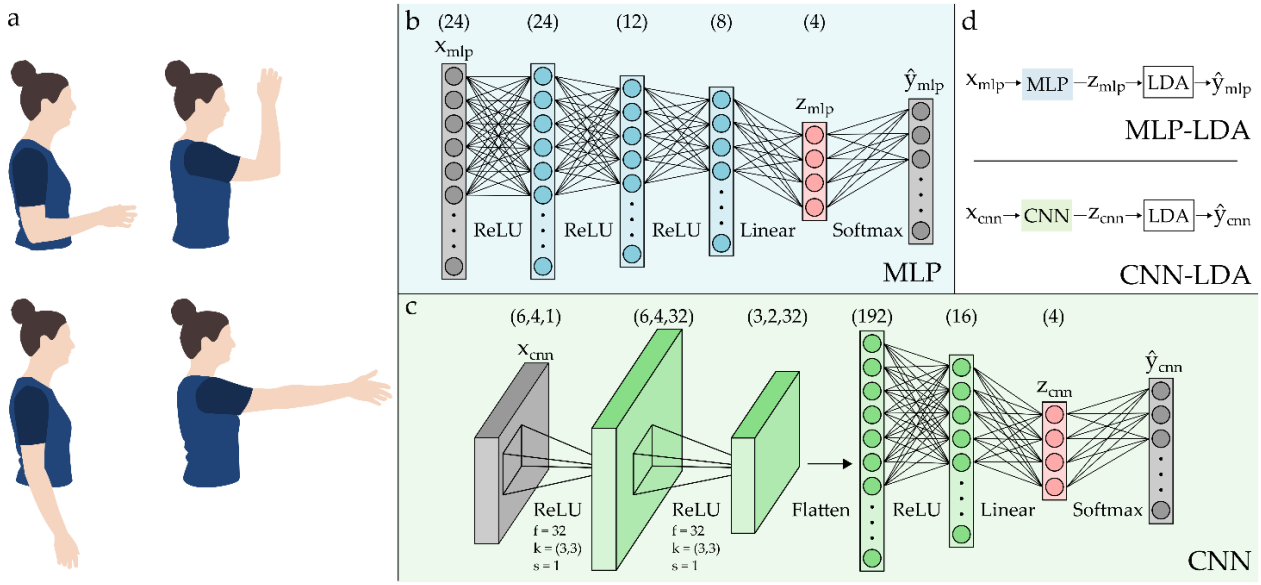

Figure 3-3 Details of training protocol and neural networks. (a) To increase test data variability, participants completed wrist and hand gestures in four arm positions. Network architectures for the (b) MLP and (c) CNN models, where f = number of filters, k = kernel size, and s = stride length. (d) The MLP-LDA and CNN-LDA models used neural networks to compute EMG latent features that were classified by LDA classifiers. The MLP-LDA+ and CNN-LDA+ utilized the same architecture but were trained with a different data set.

3.3.6 Evaluation

To evaluate control performance and robustness, we calculated the offline accuracies of each gesture classification strategies on clean and noisy test data. Since it was impractical and challenging to introduce interface noise in a controlled manner during data collection, we constructed noisy test data offline by fusing the original test raw signals with examples from a real noise database.

Real Noise Database

The effects of four noise types were investigated in this study: broken wires, moving broken wires, contact artifacts, and loose electrodes. A database containing 25 seconds (1000 overlapping windows) of each type was collected from one ITL subject (Fig. 3-4). Since the housing of the armband prevented access to individual electrodes, this database was recorded using the wet electrode setup. Although all six channels were recorded, only the affected noisy channel was stored in the database.

To simulate the broken wire and moving broken wire conditions, one wire was cut at the connection point between the wire and the electrode. For the broken wire condition, the subject maintained a 90-degree angle at the elbow throughout data collection. For the moving broken wire condition, the subject moved their arm around freely in a workspace that contained sources of electrical noise, such as monitors and laptops. For the contact artifact condition, the electrode was tapped approximately every 200 ms. Finally, for the loose electrode condition, the electrode was peeled off and gently shifted around the surface of the subject's skin throughout signal recording.

Fusion of Test Signals and Real Noise

We constructed four noisy test sets, each containing a distinct number of noisy EMG channels (1 to 4 channels). Each noisy set started as a copy of the clean test raw signals. We then systematically superimposed pseudorandomized samples from the real noise database onto the copy, ensuring that all combinations of affected channels and noise types were equally represented. To maintain signal amplification consistency, the subject-specific channel gains were applied to the noise windows according to the channels with which they were being fused. Signals were then truncated to stay within the output range of [-5V, 5V]. Finally, we extracted the four time-domain features from the noisy test signals.

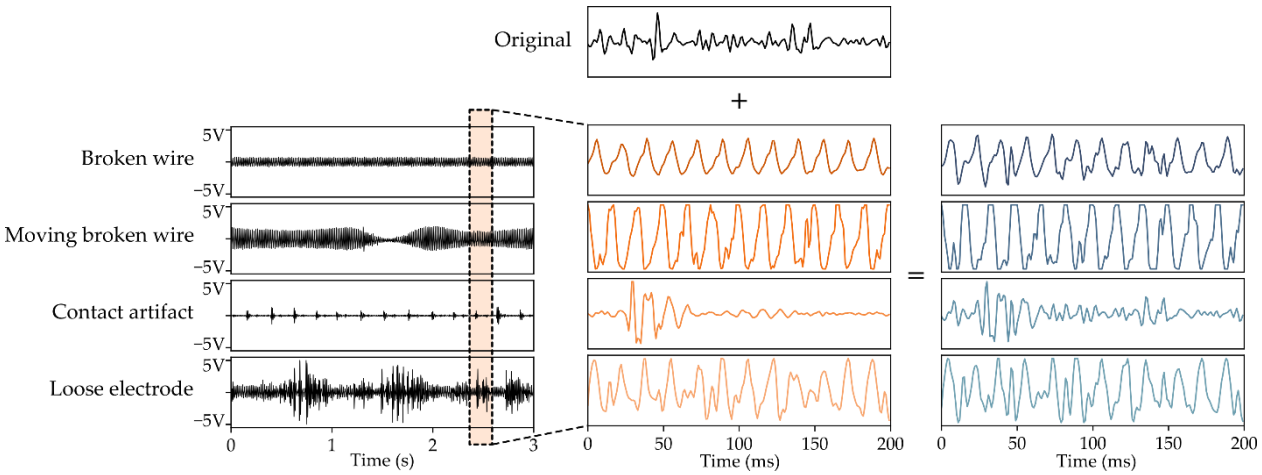

Figure 3-4. Examples from the real noise database fused with a clean EMG signal.

3.3.7 Statistical Analyses

The statistical analyses were conducted separately for ITL and AMP populations. We used linear mixed effects models to evaluate the statistical effects of each control algorithm with respect to the baseline LDA method. Initially, we fit a model with classification accuracy as the response variable, the classification strategy (LDA, LDA+, LDA-, MLP-LDA, CNN-LDA, MLP-LDA+, CNN-LDA+), number of noisy electrodes (0-4), and their interactions as fixed factors, and the subject identifier as a random factor. Statistical significance was judged based on a significance level of $\alpha = 0.05$. If the interaction factors were statistically significant, the data were separated by the number of noisy electrodes. These data subsets were used to fit five new models that each had the classification strategy as a fixed factor and subject identifier as a random factor. We used the Bonferroni method to correct for multiple comparisons.

3.4 Results

The classification accuracies of the seven classifiers and summary of the statistical models are shown in Fig. 3-5.

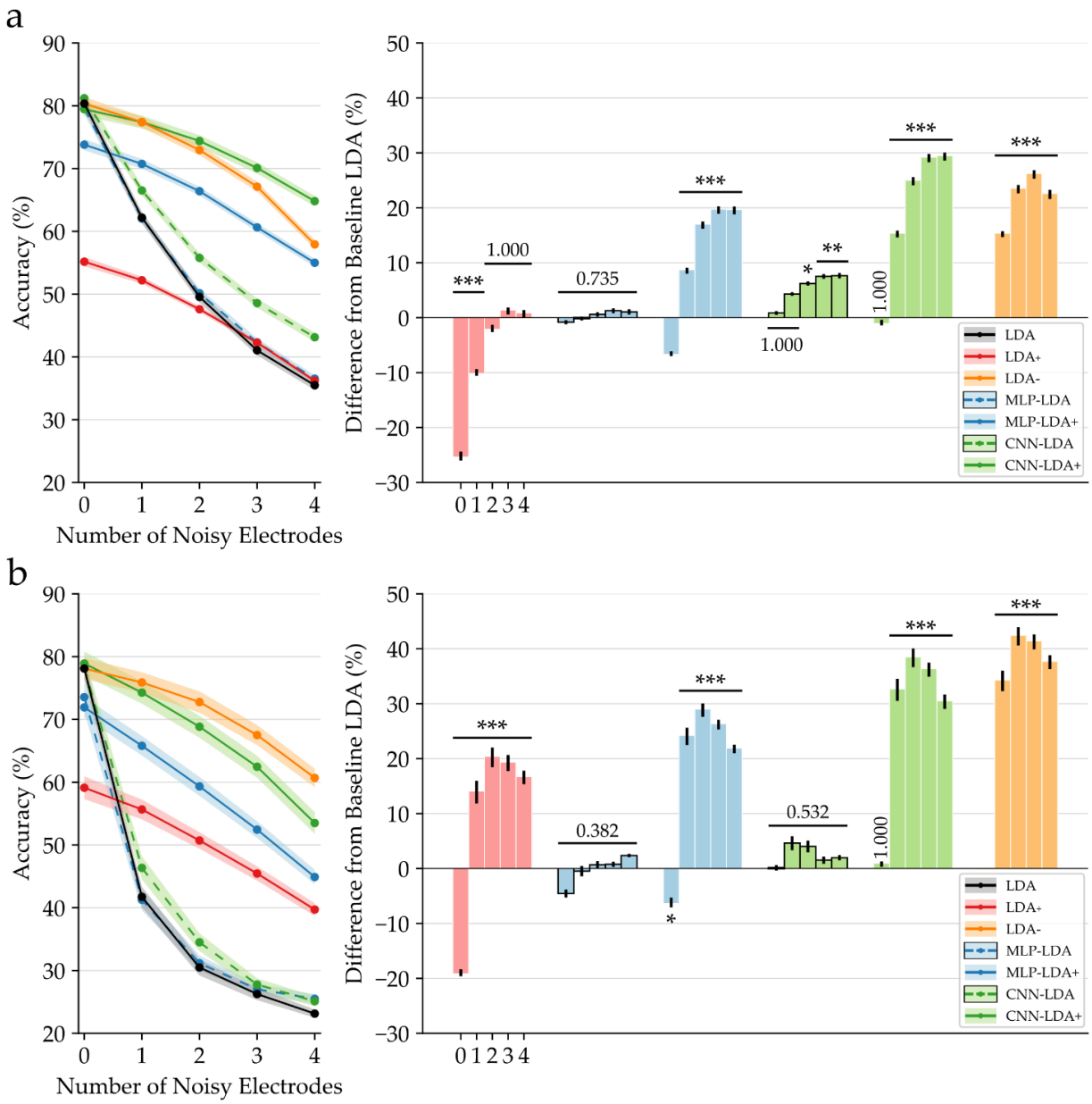

Figure 3-5. Average classification accuracies and differences from baseline LDA accuracies for (a) intact limb participants and (b) amputee participants. Note: MLP-LDA is obscured by LDA due to their similar results.

Interface noise degrades LDA accuracy

Under standard conditions, the baseline LDA classifier decoded gestures with average accuracies of 79.92 ± 1.14% and 78.10 ± 1.66% for ITL and AMP participants respectively. When noise was present in just one channel, the accuracies dropped to 61.55 ± 0.98% and 41.74 ± 1.62%, demonstrating that a minor change in input signals can have a large impact on control performance,

particularly for AMP subjects. As the number of corrupted channels increased, the accuracy continued to decrease.

Data augmentation increases noise tolerance, but often at a cost

For the AMP population, augmenting the training data set with synthetic noise increased the robustness of all classifiers. Compared to the 36.36% drop in baseline LDA accuracy between the noiseless and single channel conditions, LDA+ accuracy decreased by only 4.19%. Consequently, the LDA+ method significantly outperformed the baseline LDA method for all noisy conditions ($p < 0.001$). However, with an accuracy of 59.12 ± 1.79%, the LDA+ algorithm was also significantly worse at classifying clean signals compared to the baseline LDA algorithm ($p < 0.001$). Similarly, noise robustness improved from MLP-LDA to MLP-LDA+ and CNN-LDA to CNN-LDA+. Like the conventional LDA algorithms, using augmented training data with the MLP-LDA models also reduced classification accuracy on clean signals. In contrast, CNN-LDA+ maintained its high accuracy on clean data.

The effects of data augmentation on MLP-LDA+ and CNN-LDA+ were reiterated in the ITL population. However, the trends for the LDA+ classifier were different. We found that the LDA+ outcomes were significantly worse compared to those of the baseline LDA classifier for the noiseless and single channel noise conditions ($p < 0.001$), but not significantly different for the remaining noisy conditions.

Augmented neural network models outperform non-adaptive LDA algorithms

Generally, the neural network-aligned methods trained with the augmented data set improved overall outcomes compared to all non-adaptive LDA methods. Across all test sets, MLP-LDA+ had average accuracies of 64.84% (ITL) and 58.88% (AMP), which were 18.72% and 7.73% higher than LDA+ accuracies. MLP-LDA+ also improved classification of noisy signals by

15.77% (ITL) and 25.22% (AMP) compared to the baseline LDA classifier. However, there were statistically significant drops in accuracy on clean signals (ITL: 6.13%, AMP: 6.18%). Thus, at best, MLP-LDA+ had a 73.78 ± 0.99% accuracy for ITL subjects and a 71.92 ± 1.72% accuracy for AMP subjects.

In contrast, the CNN-LDA+ strategy significantly improved classification of noisy EMG signals ($p < 0.001$) without decreasing accuracy on clean EMG signals ($p = 1.000$). CNN-LDA+ classified normal signals with an accuracy of 80.25 ± 1.21% (ITL) and 78.91 ± 1.89% (AMP), exhibiting the best performances across all five gesture classification strategies. Unsurprisingly, these accuracies decreased as noise was introduced into the system. However, CNN-LDA+ scored 65.52 ± 0.90 % (ITL) and 53.49 ± 1.71% (AMP) with four noisy channels, meaning that at its worst, it still performed better than baseline LDA did with only one noisy channel. Therefore, CNN-LDA+ was the most accurate and robust non-adaptive method.

CNN-LDA+ eliminates need for controller adaptation

Overall, we found that CNN-LDA+ and LDA- achieved similar performances. For AMP participants, LDA- slightly outperformed CNN-LDA+ on noisy data; the differences in accuracy ranged from -1.61% for single-channel noise to -7.20% for four-channel noise. For ITL subjects, CNN-LDA+ accuracies surpassed LDA-, with improvements ranging from 0.79% for single-channel noise and 6.87% for four channel noise. Although we did not statistically compare these differences, it is unlikely that they would cause significant clinical impact. Thus, CNN-LDA+ was functionally equivalent to an adaptive LDA control system with a perfect fault detector.

LDA components illustrate performance differences

To visualize the underlying mechanisms of the non-adaptive classifiers, we plotted the first three LDA components of the clean training data, clean test data, and noisy test data for subject

AMP5 (Fig. 3-6). For baseline LDA and LDA+, we used the projection matrices from the trained LDA classifiers to reduce the high-dimensional input features. For MLP-LDA+ and CNN-LDA+, we aligned the input feature data to their latent manifolds through the MLP and CNN encoders before applying the projection matrices from their corresponding LDA classifiers. Since the performances of MLP-LDA and CNN-LDA were similar to LDA, we omitted them from the latent space visualization.

CNN-LDA+ was the only classifier that maintained separable clusters across all three data sets, explaining why it was able to effectively classify both clean and noisy signals. In contrast, LDA and MLP-LDA+ clusters lost their separability when noise was introduced, thus depicting their lower noise tolerance. The LDA+ clean test set clusters did not match the clean training set clusters; therefore, the decision boundaries computed from the training set were not able to delineate the gestures in clean test data set.

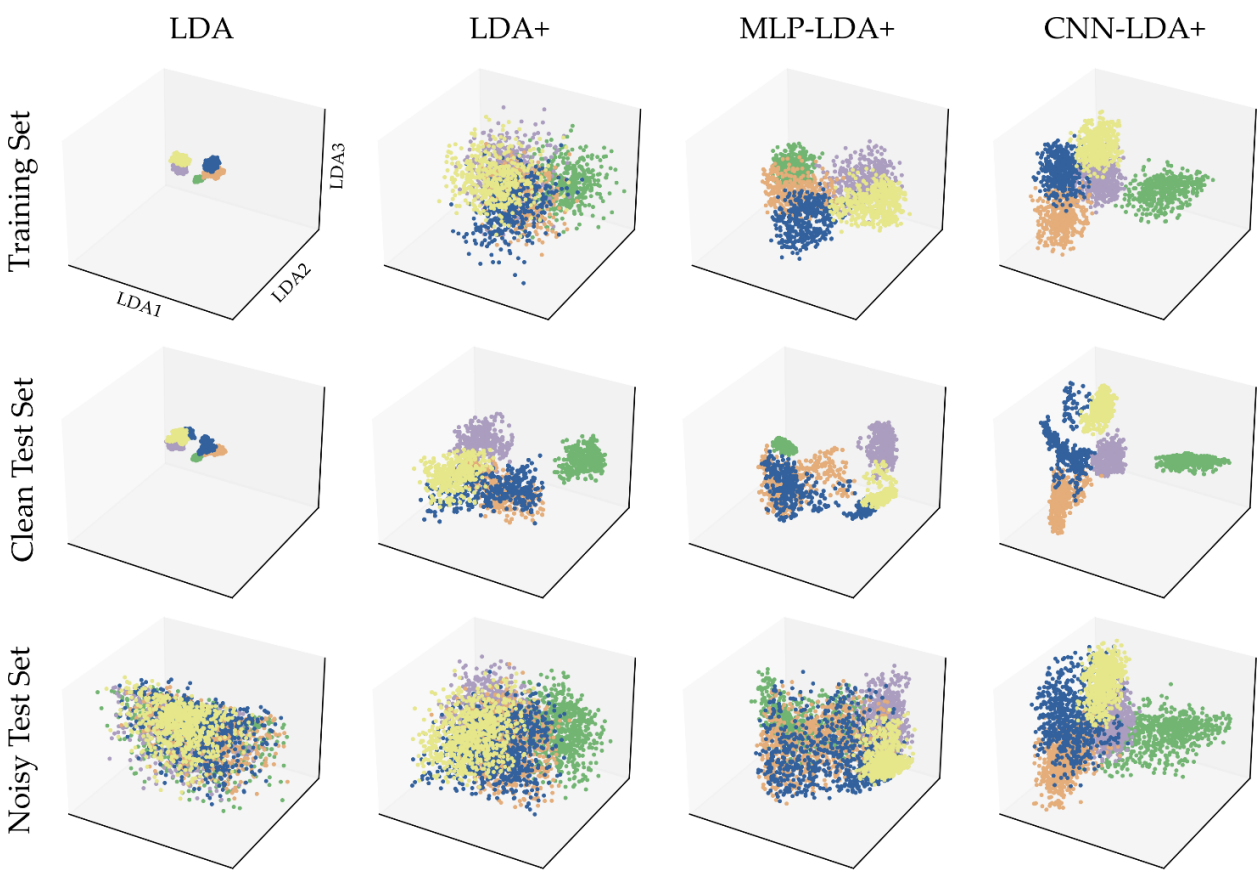

Figure 3-6. Three-dimensional latent representations of AMP5's training and test data sets using LDA components from LDA, LDA+, MLP-LDA+, and CNN-LDA+.

Neural network methods require longer processing times

In Table 3-3, we show the processing times of each control method to assess their practicality in clinical settings. Notably, the initial training processing times for the augmented neural network methods, which included data augmentation time, were substantially longer than those for the traditional LDA methods. Likewise, the classification processing times were longer for the neural network methods. However, at ~5 ms, this was still shorter than the EMG window sampling increments (25 ms), indicating that a real-time implementation is plausible.

Table 3-3. Controller Processing Times

	Training Processing Time	Classification Processing Time
LDA	0.0259 s ± 1.37 ms	0.00659 ms ± 64.4 ns
LDA+	1.22 s ± 48.6 ms	0.00665 ms ± 68.8 ns
LDA-	0.0259 s ± 1.37 ms	0.328 ms ± 4.02 µs
MLP-LDA	3.01 s ± 1.34 s	5.54 ms ± 38.1 µs
MLP-LDA+	21.4 s ± 1.83 s	
CNN-LDA	3.09 s ± 1.07 s	5.75 ms ± 38.7 µs
CNN-LDA+	22.6 s ± 1.03 s	

3.5 Discussion

Clinical PR-based myoelectric control systems often encounter error-producing signal noise stemming from EMG interface instabilities. Previous works have reduced these errors with additional processing steps such as signal filtering, noise detection, and controller recalibration or adaptation. To our knowledge, this is the first study that presents a PR control strategy that is inherently resistant to various types of multichannel interface noise. Using examples of real interface noise, we evaluated the reliability of LDA-based classifiers that employed data augmentation and deep learning techniques.

In general, we found that the adverse effects of interface noise were more severe in the amputee population than the intact limb population. A potential reason is that the intact limb population training data may have had greater variability due to the use of dry electrodes. Since dry electrodes have higher impedances than wet electrodes, their EMG signals may contain more signal noise. Hence, the intact limb population's training data would have resembled the test data more closely, resulting in more robust classifiers. Additionally, intact limb muscle contractions often have larger amplitudes compared to residual limb contractions. Therefore, noisy intact limb EMG signals have higher signal-to-noise ratios (SNR), leading to better classification rates.

The high SNR in intact limb participants may also explain why augmenting the training data set with artificial noise did not improve overall outcomes for the augmented LDA classifier. The augmented data set contained various levels of signal corruption that spanned the output range of the EMG channels. Consequently, it had a wide range of SNRs as opposed to the relatively high SNRs of the noisy test set. Thus, the augmented LDA classifier was trained on distributions that were not representative of the testing data, causing it to perform poorly.

Our results also showed that manifold alignment facilitated by neural networks trained with an augmented data set produced impactful performance gains. This suggests that nonlinear transformations are crucial to extracting useful discriminative structures within EMG signals. Furthermore, the CNN methods outperformed the MLP methods, emphasizing that the spatially local patterns between channels are valuable and should be preserved and leveraged for better control. The poor performances of the non-augmented neural networks and augmented LDA+ show that data augmentation or deep learning alone are not enough to improve performance in a meaningful way. Thus, they should be used in tandem to achieve robust classifiers.

For the intact limb population, the CNN-LDA+ method was more accurate than the adaptive LDA- method. This exemplifies an important advantage of the CNN-LDA+ method: it can retain discriminative features from the noisy channels that would normally be discarded by recalibration methods such as LDA-. In addition to the physiological differences between subject populations, the dry electrode setup may also explain why CNN-LDA+ was more effective for intact limb participants. Our dimensionality reduction approach exploits EMG signal redundancies across channels, meaning that diffuse EMG measurements are more informative than localized measurements. Hence, signals obtained by the dry electrodes, which had slightly larger surface areas, may have facilitated latent space disentanglement. Notably, CNN-LDA+ was able to classify

signals with real interface noise even though the augmented training data set only contained synthetic noise, highlighting the model's generalizability.

Ultimately, the CNN-LDA+ classification method is an attractive solution to the interface noise problem, as it was the only non-adaptive method that improved noise tolerance without reducing accuracy on normal signals. In reality, disturbances would most likely occur in one or two EMG channels. For these cases, CNN-LDA+ obtained classification accuracies (ITL: >75.19%, AMP: >68.83%) that suggest acceptable real-time control; however, real-time functional tests are still needed to verify the efficacy of the classification strategy (Young et al., 2011).

The clinical implications of the CNN-LDA+ classification strategy are that patients would be able to maintain good control of their prosthesis across common scenarios that produce intermittent noise (eg. electrode liftoff, contact artifacts) or continuous noise (eg. broken wires). Moreover, it does not require a noise detector, which can be inaccurate, or recalibration, which adds a processing step. This method is highly beneficial for amputee users, as their low EMG SNR renders their clinical LDA systems unusable in the presence of noise.

However, there are some practical limitations to the CNN-LDA+ classifier. For example, its training and execution processing time is slower than traditional LDA methods. Also, the memory requirements of the model may hinder its implementation on a prosthesis microcontroller. Lastly, it is difficult to quickly retrain a black box model such as the CNN-LDA+. If the user wanted to recalibrate a single gesture, the entire backpropagation process would have to be repeated.

3.5.1 Limitations and Future Work

The main limitation of this study is that the controllers were evaluated offline. The relationship between offline classification accuracy and real-time control performance is not well-

defined; thus, the practical significance of our findings is limited. Real-time experiments should be conducted to investigate the effects of user adaptation and provide a more realistic evaluation of prosthesis controllability. Additionally, while commercial systems typically include eight EMG channels, the experimental setup only used six EMG channels, meaning that each noisy signal had a greater influence on classification accuracy. Thus, we expect the performance to be better with clinical setups that have eight EMG channels.

Four amputee subjects only completed enough gestures for a 2DOF controller instead of 3DOF controller. These algorithms were also limited to sequential control but would be more impactful if extended to simultaneous control. To facilitate practical implementation of the CNN-LDA+, we plan to further minimize the network architecture and adjust hyperparameters to balance controller performance and processing time. In addition to conducting real-time experiments with a physical prosthesis, it would be useful to investigate the controller's robustness to donning/doffing and its long-term stability.

3.6 Conclusion

In this study, we investigated the use of training data augmentation and deep learning models to achieve reliable myoelectric control of hand and wrist gestures. Our results showed that when trained with a data set that includes synthetic noise, a CNN-based classifier can learn latent representations of muscle patterns that are linearly separable and robust to multichannel interface noise. The performance gains exhibited by the CNN-LDA+ are a significant improvement on clinical myoelectric PR controllers and motivate further development.

4 Fast generative replay-based adaptation of deep convolutional networks decreases myoelectric controller recalibration frequency

4.1 Introduction

Despite decades of advancements, robust and intuitive control of myoelectric prostheses remains challenging. Clinically available control systems can enable reliable control in stationary environments where EMG inputs are repeatable, but deteriorate when there is a significant mismatch between training and testing signal distributions.

This mismatch is a well-known machine learning problem called dataset shift, which can be categorized into two types (Moreno-Torres et al., 2012; Webb et al., 2016). The first type is covariate shift, or a change between the training and test distributions of the input signals P(X) while the class conditional probability P(Y|X) remains constant. Covariate shift often occurs when the prosthesis is trained with data that are not fully representative of data from real-world scenarios, also known as sample selection bias. For example, previous work showed that it is difficult to control a prosthesis in a limb position that was not used during training data collection (Campbell, Phinyomark, et al., 2020; E. J. Scheme et al., 2010). Another source of covariate shift is interface noise that can be caused by fluctuations in residual limb volume, broken wires or electrodes, intermittent electrode lift-off, and more. The second type of dataset shift, concept drift, is defined as a change in the conditional probability between the input and output signals P(Y|X) while the distribution of the input signals P(X) remains constant. Examples of concept drift are less common than covariate shift but may occur as users adapt their muscle patterns over time.

When dataset shift reduces classification accuracy to the point where control becomes ineffective, users can choose to recalibrate their prosthesis. This is often the case for clinically

available control algorithms, which are not yet robust to common sources of dataset shift. As such, they require frequent recalibrations to maintain adequate control (Levi Hargrove et al., 2018).

Recently, deep learning techniques have been developed to improve myoelectric control robustness to dataset shift (Ameri et al., 2019, 2020; Campbell, Member, et al., 2020; ur Rehman et al., 2018; Wang et al., n.d.; Wu et al., 2020). Despite their positive results in experimental settings, they are not immune to all sources of dataset shift and will likely need recalibration from time to time. Therefore, an effective and efficient classifier recalibration strategy is necessary to facilitate clinical implementation of deep learning control algorithms.

There are several approaches we can take to recalibrate a deep neural network, not all of them are clinically feasible. The most conservative method is to store all previously collected data and train the model from scratch using the accumulated data. This ensures that the network learns from the most comprehensive set of examples. However, this would require memory, time, and computational power that cannot be afforded by prosthesis hardware. We could also train a model from scratch using only the most recently collected data set. However, the model could suffer from catastrophic forgetting, where it loses all previously learned information (van de Ven & Tolias, 2019). Depending on the size and quality of the new data set, it may also take a long time to train or be prone to overfitting.

Finally, we can use ideas from continual learning, a branch of machine learning that seeks to sequentially update models with new information without losing previously learned relationships (Aljundi, 2019; Delange et al., 2021). The most basic continual learning concept is finetuning, where a pretrained model (or the model trained on previous data) is updated for several iterations with the new training data. This relies on the same mechanism as early stopping to avoid

overfitting to the new data and erasing previous information. One approach of continual learning that has found success is replay-based continual learning, where some form of past data is fed to the network while it is being trained (Pellegrini et al., 2020; Rolnick et al., 2019). If there are no memory constraints, we could simply store all previous data sets. A more efficient method is generative replay, where we learn the distributions of the data and use that information to generate new examples during training time (Shin et al., 2013; Ven & Tolias, 2019).

The objective of this study was to quantify dataset shift during long-term prosthesis usage and use continual learning concepts to develop clinically feasible retraining techniques for deep convolutional neural network-based myoelectric controllers. We evaluated these techniques by simulating a sequential recalibration process over six months of EMG data from seven individuals with upper limb loss.

4.2 Methods

In this post hoc data analysis study, we quantified dataset shift and evaluated two classifiers and five recalibration techniques across six months of EMG control signals. This was conducted using data collected from seven individuals with unilateral below elbow amputations in their home setting (Table 4-1). Although eight subjects participated in the home trials, one subject was omitted from this work because they did not have sufficient data. Eight weeks into the study, participants underwent targeted muscle reinnervation (TMR) surgery. After recovery, data collection resumed for 20 more weeks. This study was approved by Northwestern University's and Walter Reed National Military Center's Institutional Review Boards. All participants gave written informed consent.

4.2.1 Home Trial Protocol

At the start of the study, a certified prosthetist fitted each participant with an OSSUR i-limb Ultra Revolution multi-articulating hand and a passive wrist rotator. The prosthesis was controlled with surface EMG signals through a custom version of the Coapt COMPLETE CONTROL system that enabled data logging. Eight grip types were available to be configured, including lateral, power, thumb precision pinch opened and closed, thumb 3 jaw chuck opened and closed, standard 3 jaw chuck opened and closed, and index point. Muscle signals were collected through eight pairs of dry stainless-steel electrodes that were embedded in a silicone liner and worn under a socket. The electrode locations were empirically determined by the prosthetist to enable effective direct and pattern recognition control.

After participating in occupational therapy to learn how to effectively use their prosthesis to perform activities of daily living, each participant completed two 8-week (one pre-TMR and one post-TMR) and one 12-week (post-TMR) home trials with their device. During these trials, participants were able to recalibrate the device's pattern recognition control algorithm at will by using a prothesis guided training protocol. In each of these recalibration sessions, users followed a pre-programmed sequence of gestures while the embedded system measured and recorded EMG signals using 200 ms windows in 25 ms increments. The calibration sequence always began with the 'rest' state, followed by two repetitions of each grip type interleaved with the 'hand open' gesture between each grip repetition. Each gesture repetition was held for 3 s. Thus, if there were 3 pre-selected grip types, the calibration session contained 6s of each grip type, 3s of rest, and 15s of the 'hand open' gesture. The system automatically labeled the grip types based on the pre-programmed gesture sequence and extracted time-domain and autoregressive features from each data window. These features and labels were used to train a linear discriminant analysis (LDA)

classifier that provided pattern recognition control for their myoelectric prosthesis. The raw EMG signals, gesture labels, and time of each recalibration session were recorded, thus providing us with snapshots of the EMG data distributions over time.

Table 4-1. Amputee Population Demographics

Subject	Gender	Age	Years Post Amputation	Etiology	Prescribed Prosthesis	
					Terminal Device	Myoelectric Control
TR1	Male	30	2	Left Trauma	Bebionic Hand	Coapt Pattern Recognition
TR2	Male	40	3	Left Trauma	Bebionic Hand, Motion Control ETD	Direct Control
TR3	Female	49	12	Right Trauma	Bebionic Hand, Motion Control ETD	Coapt Pattern Recognition
TR4	Male	32	1	Right Trauma	Michelangelo Hand	Direct Control
TR5	Male	42	0.8	Left Trauma	Sensor speed hand, ETD	Direct Control
TR6	Male	58	1	Right Trauma	ETD	Direct Control
TR7	Male	53	12	Right Trauma	i-Limb, Sensor speed hand	Direct Control

ETD: electric terminal device

4.2.2 Data Cleaning

EMG signal quality varied across the recalibration sessions. Some sessions contained EMG interface noise, which may have been caused by wire breakage, intermittent electrode lift-off, and residual limb volume changes. Since we wanted to evaluate the classifier robustness to noise, we did not remove most of these noisy recalibration sessions. However, sessions in which more than two channels were visibly affected (eg. disconnected ground electrode) were omitted from data analysis. Additionally, since the at-home recalibration sessions were unsupervised by a researcher or clinician, several sessions contained user timing errors that resulted in inaccurate gesture labels. Sessions that contained obvious user errors (eg. no signals during gestures, contractions that

overlapped across multiple gestures) were omitted from data analysis. Summaries of the data cleaning results are shown in Tables 4-2 and 4-3.

To reduce the prevalence of user timing issues in the remaining data sets, we applied a simple data relabeling scheme. We used the mean absolute value (MAV) of the signals corresponding to the 'rest' gesture as a threshold for the remaining gestures. If a data window had an MAV that fell below 1.1x of the threshold, it would be relabeled as the 'rest' class.

Table 4-2. Data cleaning for pre-TMR recalibration sets

Subject	Total	Noisy sets discarded	User error sets discarded	Sets used
TR1	12	0	6	**6**
TR2	12	0	0	**12**
TR3	22	0	0	**22**
TR4	91	9	3	**79**
TR5	30	5	0	**25**
TR6	39	17	1	**21**
TR7	11	1	1	**9**

Table 4-3. Data cleaning for post-TMR recalibration sets

Subject	Total	Noisy sets discarded	User error sets discarded	Sets used
TR1	12	1	1	**10**
TR2	8	0	0	**8**
TR3	28	0	0	**28**
TR4	42	0	1	**41**
TR5	23	0	0	**23**
TR6	23	0	1	**22**
TR7	25	2	3	**20**

4.2.3 Post-hoc Recalibration Protocol

We used the recalibration sessions chronologically to evaluate the robustness of two classification strategies and their recalibration methods over time. The first strategy was a traditional LDA classifier, representing clinically available control systems such as the one used

during the home trials. The second strategy was a CNN classifier adapted from Chapter 3, where it was shown to be effective at increasing noise tolerance. Both classifiers were trained with four time-domain features (MAV, waveform length, slope sign changes, and zero crossings) and six autoregressive coefficients.

The pre-TMR and post-TMR data blocks were separated for this analysis. For each, we first trained baseline classifiers LDA_0 and CNN_0 on an initial training set (X_0, Y_0), where X_0 denotes the inputs and Y_0 denotes the gesture labels. The initial data set was composed of EMG signals from the first two recalibration sessions of each data block. Half of the data set was used for training and the other half was used for testing. We performed five-fold cross validation, with each fold using a unique training/test data split.

The network architecture of the CNN classifier is shown in Fig. 4-1. We trained the baseline classifier for 30 epochs using the Adam optimizer with a learning rate of 0.001 and batch size of 128. Additionally, data augmentation was performed on the training data set to increase training data variability. Further details about this classifier and data augmentation technique can be found in Chapter 3.

To evaluate recalibration methods, we started with the baseline classifiers and used them to classify the remaining data sets sequentially. If the classification accuracy fell below a certain threshold, we used our proposed methods to recalibrate the classifier. The full recalibration protocol for a participant with N recalibration sessions was as follows:

1. We trained a baseline classifier on the initial training set (X_0, Y_0).
2. We tested the classifier on next data set (X_1, Y_1).

3. If the classification accuracy was less than 70%, we recalibrated the classifier with half of (X_1, Y_1) and validated with the other half. As with the initial training data set, we performed five-fold cross validation, using a different training/test data split for each fold. The recalibration threshold of 70% approximated the classification accuracy at which users decided to recalibrate their device. We calculated this by testing each data set (X_n, Y_n) using an LDA classifier trained on the previous data set (X_{n-1}, Y_{n-1}). We took the median classification accuracy across all *N* calibration sets and averaged this value across participants. The threshold we obtained corroborated previous work that found that offline classification accuracies under 65% led to poor real-time control.
4. If the validation accuracy was less than 70%, this counted as a failed recalibration and we reverted the classifier to its previous weights, ie. LDA_{n-1} or CNN_{n-1}.
5. We repeated steps 2-4, incrementing through all *N* recalibration sets. Throughout this process, we tallied the number of successful and failed recalibrations to quantify the efficacy of our recalibration methods.

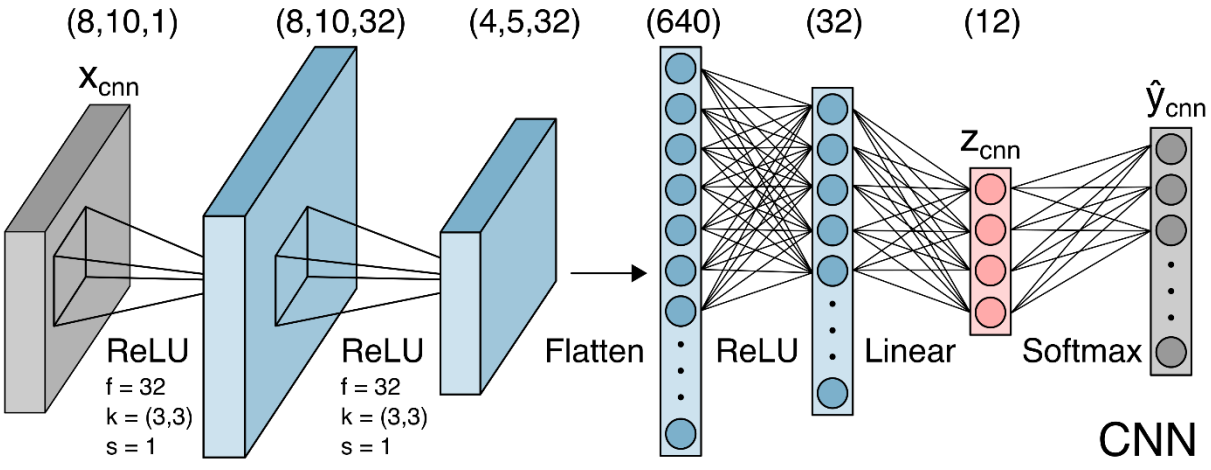

Figure 4-1. Network architecture for the CNN classifier.

We used this procedure to test five recalibration techniques, described below.

1. Recalibrated LDA (LDA) – In this method, an LDA classifier was trained on the current data set only. This is the typical recalibration method for pattern recognition-based prosthesis.
2. Adapted LDA (a-LDA) – This method stored the mean and covariance matrices of previous training data ($\mu_{0:n-1}, \Sigma_{0:n-1}$) and the cumulative number of training examples per class. During recalibration, we update these values using the formula shown below, where $S_{i,n}$ denotes the number of examples in class i from dataset (X_n, Y_n), and $S_{i,0:n-1}$ denotes the accumulated number of examples in class i from datasets (X_0, Y_0) to (X_{n-1}, Y_{n-1}). The new mean and covariance matrices ($\mu_{0:n}, \Sigma_{0:n}$) are then used to calculate the LDA classifier weights.

$$\alpha = \frac{S_{i,0:n-1}}{S_{i,0:n-1} + S_{i,n}}$$

$$\mu_{i,0:n} = \alpha\mu_{i,0:n-1} + (1-\alpha)\mu_{i,n}$$

$$\Sigma_{i,n} = \left(X_{i,n} - \mu_{i,0:n-1}\right)^T \left(X_{i,n} - \mu_{i,0:n}\right)$$

$$\Sigma_{i,0:n} = \alpha\Sigma_{i,0:n-1} + (1-\alpha)\Sigma_{i,n}$$

3. Recalibrated CNN (CNN) – We reinitialized the parameters of the CNN classifier and retrained it with the new data set after data augmentation. The Adam optimizer was used to train the network for 30 epochs with a batch size of 128 and learning rate of 0.001, matching the hyperparameters of the initial training procedure for CNN_0.
4. Finetuned CNN (ft-CNN) – The most recent CNN classifier was trained on new data set after data augmentation for 5 epochs with a batch size of 128 and learning rate of 0.001.

5. Generative replay CNN (gr-CNN) – Similar to the a-LDA method, the gr-CNN stored running mean and standard deviation matrices that parameterized previous class distributions, $(\mu_{i,0:n-1},\ \sigma_{i,0:n-1})$. It also stored the number of training examples for each class. During recalibration, we sampled from multivariate Gaussian distributions $\mathcal{N}(\mu_{i,0:n-1},\ \sigma_{i,0:n-1})$ to generate training data (X_{gen}, Y_{gen}) that resembled previously seen training examples. We then concatenated (X_{gen}, Y_{gen}) with an augmented (X_n, Y_n) and used the combined training data to update the most recent CNN by training for 5 epochs with a batch size of 128 and learning rate of 0.001.

4.2.4 Outcome Measures

Dataset shift

To quantify dataset shift over time and assess its impact on classification accuracy, we computed the squared Hellinger distance between the initial training distribution and subsequent test distributions (Beran, 1977; Goldenberg & Webb, 2019; Nilsson et al., 2017). Specifically, we measured the distance between the total input distributions $P(X_0)$ and $P(X_n)$ and the distance between the class distributions $P(X_0|Y)$ and $P(X_n|Y)$, where X_0 denotes the inputs from the first calibration set and X_n denotes the inputs from the n^{th} recalibration set. The input feature vectors were projected to the trained LDA_0 classifier space to reduce the dimensions prior to calculating the Hellinger distance. The Hellinger distance is a bounded metric between 0 and 1, where 0 describes indiscernible distributions. This makes it easy to normalize data across multiple subjects. We assume that the data are multivariate normal distributions. Thus, the Hellinger distance between $P \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $Q \sim \mathcal{N}(\mu_2, \Sigma_2)$ can be calculated as follows:

$$\Sigma = \frac{\Sigma_1 + \Sigma_2}{2}$$

$$H^2(P,Q) = 1 - \frac{(det\Sigma_1)^{\frac{1}{4}}(det\Sigma_2)^{\frac{1}{4}}}{(det\Sigma)^{\frac{1}{2}}} \exp\left\{-\frac{1}{8}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2)\right\}$$

Recalibration performance

We introduced two metrics, recalibration frequency and recalibration efficacy, to evaluate the performance of each recalibration method. Recalibration frequency was defined as the number of total attempted recalibrations as a percentage of the total number of recalibration sessions for each participant. An effective classification and recalibration strategy would lead to improved robustness and therefore reduce the need for recalibrations. Thus, a lower recalibration frequency was desirable. Recalibration efficacy measured the number of successful recalibrations as a percentage of the total number of attempted recalibrations for each participant. In the real world, failed recalibrations would mean that the user would have to repeat the calibration session, adding further inconvenience. An effective strategy should maximize the proportion of successful calibrations to failed calibrations, leading to a high recalibration efficacy score.

Catastrophic forgetting

To further evaluate how well the recalibration techniques could learn new information without discarding previous knowledge, we quantified catastrophic forgetting using three metrics, $\Omega_{base}, \Omega_{all}$, and Ω_{new} (Kemker et al., 2017). Ω_{base} described how well the classifier could classify the initial training set over time, Ω_{all} described how well the classifier could classify all previously seen training data, and Ω_{new} described how well the classifier could classify the most recent training data. The formulas for these metrics are shown below. $\alpha_{base,n}$ represents the accuracy of

the n[th] classifier on the initial data set, $\alpha_{all,n}$ represents the accuracy of the n[th] classifier on all previous data sets, $\alpha_{new,n}$ represents the accuracy of the n[th] classifier on the n[th] data set, and α_{ideal} was the accuracy of the baseline classifier on the initial data set and served as a normalization factor.

$$\Omega_{base} = \frac{1}{N-1}\sum_{n=2}^{N}\frac{\alpha_{base,n}}{\alpha_{ideal}}$$

$$\Omega_{all} = \frac{1}{N-1}\sum_{n=2}^{N}\frac{\alpha_{all,n}}{\alpha_{ideal}}$$

$$\Omega_{new} = \frac{1}{N-1}\sum_{n=2}^{N}\alpha_{new,n}$$

Recalibration processing time

To maintain clinical feasibility, the processing time for prosthesis recalibration should be minimal. Hence, we recorded the processing time for the initial training of the baseline classifier and each attempted recalibration for all recalibration techniques. All data analysis was conducted on a Windows 10 laptop computer with 16GB RAM, an Intel Core i7-9850H CPU at 2.60GHz, and a 4GB NVIDIA Quadro T1000 GPU.

4.2.5 Statistical Analyses

We used a linear mixed effects model to statistically compare the differences in recalibration performance and catastrophic forgetting metrics. We then conducted post hoc pairwise comparisons using paired t-tests with a significance level of $\alpha = 0.05$. To correct for multiple comparisons, p-values were adjusted using the Holm-Bonferroni method.

4.3 Results

We measured data shift magnitude using Hellinger distances (Fig. 4-2), compared recalibration frequency and catastrophic forgetting metrics across recalibration techniques (Fig. 4-4 and Fig. 4-5), and quantified processing times for each method (Table 4-4).

Quantifying dataset shift magnitude

We found that the Hellinger distances between the training and test distributions were negatively correlated with classification accuracy (Fig. 4-2). Furthermore, these distances increased over time. Together, these observations implied that the degradation of classification accuracy over time may have been rooted in dataset shift.

Specifically, when the distance between $P(X_0)$ and $P(X_n)$ exceeded 0.6, the corresponding classification accuracies were predominantly under 50%. These larger distances were also more often observed in the later stages of the home trials. Although the distance between class distributions $P(X_o|Y)$ and $P(X_n|Y)$ were also correlated with accuracy, it spanned half the range of the $P(X)$ distance. Therefore, the class distributions did not shift as much as the input distributions. We also found a strong correlation between the two Hellinger distances, suggesting that their changes were not independent.

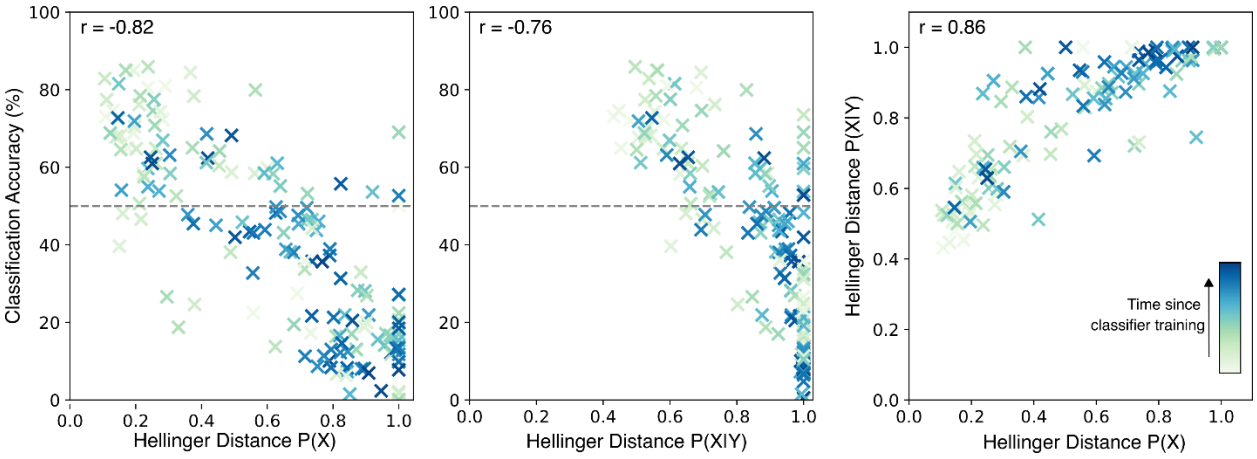

Figure 4-2. Hellinger distances between training and test distributions exhibited strong negative correlations with classification accuracy. As the time between training and test data distributions increased, depicted by the darkness of the markers, the Hellinger distance also increased.

Comparison of baseline classifiers

Fig. Figure 4-3 depicts the trajectory of classification accuracy over time for the two baseline classifiers. Qualitatively, the CNN_0 appeared to be more consistent than the LDA_0, exemplified by the narrower range of accuracies, particularly with TR1, TR3, and TR5. The average range for the LDA_0 classifier was 63.56 ± 2.35% (SE) while the range for the CNN_0 classifier was 53.80 ± 2.07%. The difference between these values was not statistically significant ($p = 0.100$). However, we found a statistically significant difference between their average accuracies, which was 66.38 ± 1.24% for the LDA_0 and 69.50 ± 1.32% for the CNN_0. Thus, before any recalibration paradigms were applied, the baseline CNN classifier outperformed the baseline LDA classifier.

Recalibration metrics

Although the number of recalibrations varied across subjects, the trends between recalibration methods were consistent (Fig. 4-4 and Fig. 4-5). Unsurprisingly, the LDA classifier

required the most recalibrations across all participants, with an average recalibration frequency of 51.17 ± 2.57%. The differences between this value and the recalibration frequencies for the a-LDA, ft-CNN, and gr-CNN were statistically significant ($p < 0.05$). Although the LDA classifier had more recalibrations than the CNN (41.48 ± 2.45%), the difference was not statistically significant ($p = 0.077$).

Using the adaptive recalibration method for the LDA classifier significantly improved outcomes. The a-LDA had fewer recalibrations than most methods (33.21 ± 1.95%), including the CNN and ft-CNN (35.03 ± 2.43%) methods, though these differences were not significant (CNN: $p = 0.096$, ft-CNN: $p = 1.000$). However, the a-LDA method also had the lowest recalibration efficacy, meaning that a larger proportion of its recalibration attempts was unsuccessful.

Overall, the gr-CNN method had the most well-rounded performance, requiring fewer recalibrations (29.20 ± 2.41%) compared to all other methods while still achieving high recalibration efficacy (83.23 ± 1.52%). Notably, the only difference between ft-CNN and gr-CNN was the addition of synthetically generated training data by sampling from normal distributions. Though simple, this method was able to produce statistically significant improvements.

The classification accuracies from the best LDA recalibration method (a-LDA) and the best CNN recalibration method (gr-CNN) were plotted in Fig. 4-3. Compared to the baseline classifiers, these recalibration methods obtained more consistent accuracies. Notably, the lower recalibration efficacy of the a-LDA model can be seen in TR3, TR4, and TR7, where the instances of failed recalibrations are apparent.

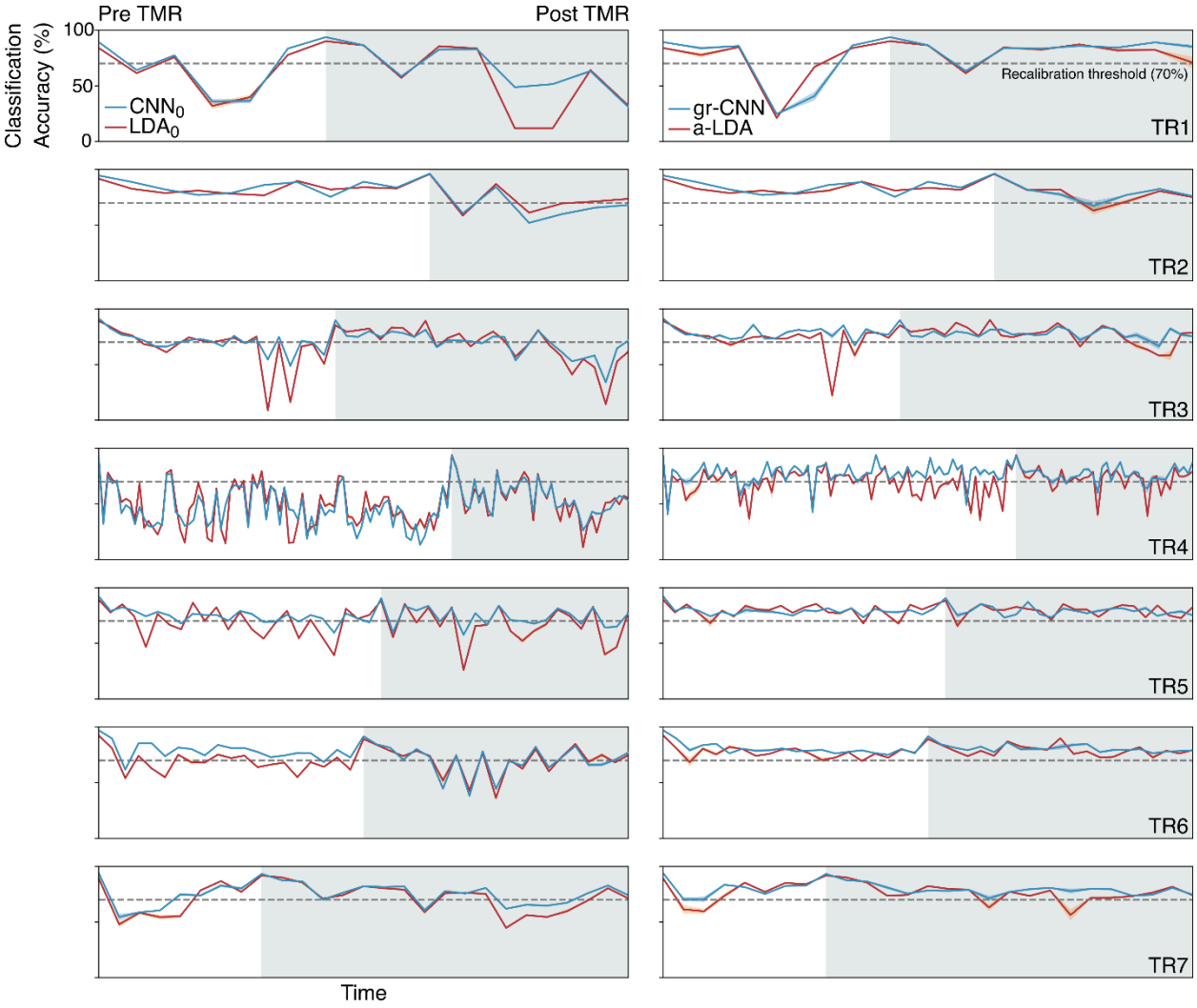

Figure 4-3. Baseline classifier accuracy over time (left) and accuracy of the best recalibration methods gr-CNN and a-LDA (right).

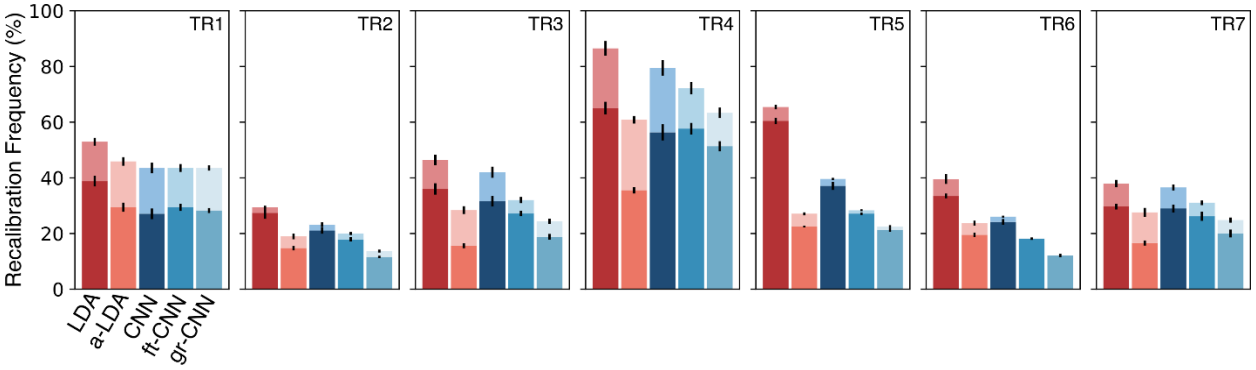

Figure 4-4. Subject specific recalibration frequencies. The top bars represent failed calibrations and the bottom bars represent successful calibrations.

Catastrophic forgetting metrics

The quantification of catastrophic forgetting elucidated the differences seen between recalibration methods (Fig. 4-5, Table 4-4). Although a-LDA had the highest Ω_{base} and Ω_{all} (94.75 ± 0.37%, 94.6 ± 0.56%), implying superior knowledge retention, it was significantly worse at learning from new training data (Ω_{new}) compared to the LDA, ft-CNN, and gr-CNN methods. This may explain its low recalibration efficacy. In contrast, the LDA and CNN strategies had the least knowledge retention, which was expected since they both reinitialized their classifiers for each recalibration, effectively wiping previously learned information. The ft-CNN and gr-CNN techniques both reduced catastrophic forgetting and were able to learn from new training data. However, the gr-CNN had significantly higher Ω_{base} and Ω_{all} measures, showing that the use of generatively sampled training data reduces catastrophic forgetting.

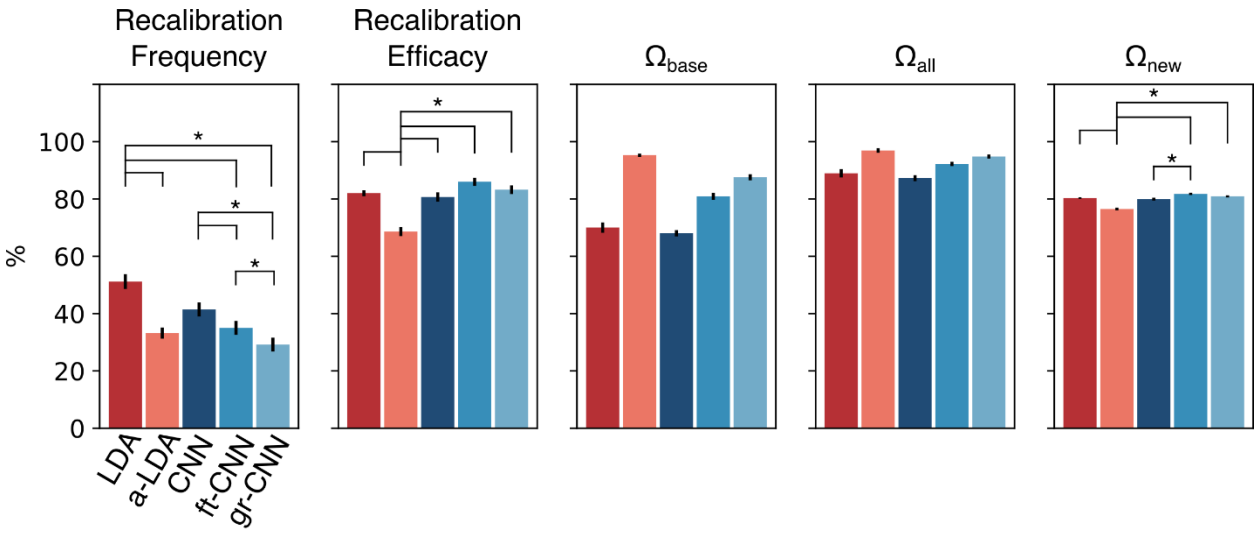

Figure 4-5. Recalibration performance and catastrophic forgetting metrics. Statistical results for Ω_{base} and Ω_{all} are presented in Table 4-4.

Table 4-4. Statistical significance of pairwise comparisons for Ω_{base} (bold) and Ω_{all}

	LDA	a-LDA	CNN	ft-CNN	gr-CNN
LDA		**	0.990	*	**
a-LDA	**		***	**	*
CNN	0.949	**		**	***
ft-CNN	*	*	**		**
gr-CNN	*	*	**	*	

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

Recalibration processing time

The classifier training and recalibration processing times, shown in Table 5, suggest that all these methods are clinically feasible. Understandably, the CNN model required the longest processing times. Even then, both the initial training and recalibration times are under 30s, which is reasonable for clinical use. However, it is unclear if training times would remain practical if these algorithms were implemented with a prosthesis microcontroller.

Table 4-5. Classifier Training and Recalibration Processing Times

	Initial Training Time	**Recalibration Time**
LDA	0.24s ± 0.03s	0.012 ± 0.0017s
a-LDA		0.014 ± 0.0014s
CNN	24.37 ± 3.68s	14.49 ± 1.75s
ft-CNN		3.04 ± 0.29s
gr-CNN		3.26 ± 0.25s

4.4 Discussion

In this study, we measured the distances between distributions across training and test data sets to quantify dataset shift magnitude over a period of six months. We then sought to improve classification robustness with recalibration techniques that prioritized efficacy and clinical feasibility. Through a simulated recalibration protocol, we quantified the efficacy of five classifier recalibration strategies.

We found that the EMG input distribution and class distributions both shifted over time and correlated with offline classification accuracy. The shifts in P(X) eliminate the possibility that the dataset shifts were caused solely by concept drift and the correlation between the P(X) and P(X|Y) distances suggest that covariate shift was also not the only cause of dataset shift. Hence, the shifts that affect long-term myoelectric control are likely a combination of changes to P(X) and P(Y|X). Furthermore, knowing that the Hellinger distance between P(Xtr) and P(Xte) is strongly correlated with classification accuracy, we may be able to design control algorithms that use this metric to inform our recalibration process. Specifically, since the measurement of P(X) does not require a gesture label, the Hellinger distance can be periodically measured with unlabeled data to detect dataset shifts and determine when to recalibrate the classifier.

Our baseline classifier results showed that the classification accuracy of a deep convolutional neural network does not fluctuate as much as that of a traditional LDA classifier over time. This control stability reiterates previous work suggesting that convolutional networks are more robust to noise. However, when dataset shift occurs and recalibration is required, the CNN quickly loses its advantages and falls behind a simpler, more robust adaptive LDA method. One potential reason for these underwhelming outcomes is that deep learning models like the CNN perform best when trained on large datasets. In this study, each recalibration sessions only consisted of two repetitions of each grip type, one repetition of rest, and five repetitions of hand open. Furthermore, the recalibration sets were split in half to obtain training and validation data. As a result, not only did the CNN erase previously learned information when it reinitialized its weights, it may have also overfitted to the small recalibration training data sets.

Fortunately, finetuning and generative replay methods can prevent catastrophic forgetting by building on top of the previously trained network and avoid overfitting by training over fewer

epochs. In addition, generating training data that describe previous training data sets further reduces catastrophic forgetting. In addition to being effective, this method is computationally inexpensive and easy to implement.

Ultimately, compared to the clinically available LDA algorithm, an adaptive LDA algorithm, finetuned CNN, and generative replay CNN can all significantly improve classifier stability over time and reduce the need for frequent recalibrations. However, the generative replay CNN offers the best balance between retaining knowledge and learning new information.

4.4.1 Limitations and Future Work

The findings of this study may have limited practical significance as all methods were implemented offline. The relationship between offline classification accuracy and real-time control is ambiguous, owing to factors such as user adaptation and sensory feedback. A study implementing these methods in real-time is necessary to provide the most realistic outcomes measures. In doing so, we must also consider other practical constraints that we were able to ignore in this study, such as embedded system limitations. In addition to classification accuracy, functional outcome measures should be evaluated to increase clinical applicability. On a related note, statistical significance may not always imply clinical significance. Thus, it is unclear if our findings would lead to clinical improvements.

Although we were able to reduce user timing errors, we could not eliminate all instances. These errors introduced a confounding variable that likely resulted in inaccurate gesture labels. The proposed recalibration methods were not designed to identify faulty training data, meaning that classifiers may have occasionally learned arbitrary relationships between EMG features and inaccurate gesture labels. Furthermore, the process of thresholding the EMG signals based on the

‘rest’ class resulted in class imbalances that may have incorrectly biased classifier training. Future work could include methods to detect user timing errors to evaluate the quality of the data set before it is used for training.

We used a classification accuracy threshold of 70% for all participants to determine when to recalibrate the classifier. Realistically, each user may have different criteria for deciding when to recalibrate their prosthesis. Indeed, we saw that the total number of calibration sets was highly variable among participants, exemplifying the importance of personal preference in prosthesis use. Conducting this study in real-time and allowing participants to voluntarily choose when to recalibrate their device would provide a realistic assessment of the proposed methods. Additionally, future work can focus on systematically identifying common reasons for recalibration and using that information to build more appropriate training tools.

4.5 Conclusion

In this study, we used EMG data collected from individuals with below-elbow amputations over the course of six months to quantify dataset shift over time and evaluate the use of LDA and CNN classifiers with five recalibration techniques. We found that dataset shift increases over time and is indeed correlated with decreasing classification accuracy. Furthermore, our results showed traditional LDA and CNN recalibration techniques are prone to catastrophic forgetting and overfitting that reduce classification accuracy. However, a simple CNN recalibration method based on generative replay can significantly reduce the need for frequent recalibrations.

5 Concluding Remarks

5.1 Summary of Contributions

The overall goal of this dissertation was to develop clinically friendly machine learning techniques to improve myoelectric pattern recognition control robustness to dataset shift. To that end, I completed three aims.

In my first aim, I looked at a commonly cited source of dataset shift: limb position. Before this study, it was generally accepted that limb position negatively affects myoelectric pattern recognition control. However, there was an overlooked gap in the literature that limited the validity of this assertion. Specifically, studies of the limb position effect on real-time control in amputees did not consider the weight of a prosthesis on the residual limb. To my knowledge, this was the first study that systematically evaluated how limb position affected real-time control in amputees when the residual limb was loaded. I found that:

- Limb position and load significantly and independently affect control performance when the controller is trained in a single position.
- Training the controller with dynamic arm movements around the workspace significantly reduces limb position effects in individuals with intact limbs.
- Limb position effects are less pronounced in individuals with upper limb loss.
- Training the controller with dynamic arm movements around the workspace eliminates limb position effects in individuals with upper limb loss.

The latter two contributions challenge long-held assumptions about the prevalence of the limb position effect. Tangentially, these results also highlight two important reminders for prosthetics research: 1) results obtained in an offline manner may not apply to online scenarios

and 2) behaviors observed in control subjects (eg. individuals with intact limbs) may not apply to individuals with limb loss.

Next, I addressed dataset shift caused by changes at the electrode-skin interface. The difficulty with interface noise is that it occurs unpredictably and when it does, even a small amount (eg. single channel noise) can render the control algorithm useless. Previous solutions to this problem included signal denoisers and adaptive classifiers. However, there were no existing classification strategies that were inherently robust to noise. To tackle this problem, I designed deep neural networks to project EMG feature inputs into a low-dimensional latent space that can be shared by noisy and clean signals. My main technical contributions and findings were:

- A data augmentation protocol to increase training data variability and improve classifier robustness.
- A deep convolutional neural network classifier that is inherently robust to noise. This classifier achieves this by learning noise-invariant features shared by clean and noisy EMG signals.
- Data augmentation generally improves classification robustness but may also reduce accuracy on clean signals depending on the classifier
- CNNs trained with augmented data can accurately classify both clean and noisy signals and performs as well as a two-stage adaptive LDA algorithm.

In my final aim, I investigated dataset shift over long durations and attempted to mitigate its negative effects with CNNs and various recalibration methods. To facilitate clinical implementation of deep learning-based classifiers, efficient recalibration methods needed to be developed. Using continual learning concepts, I created simple recalibration methods that can balance effectiveness and efficiency. My main technical contribution and findings were:

- A generative replay CNN adaptation method that can learn new information without losing knowledge from previous data.
- Dataset shift measured by the Hellinger distance increases over time and is strongly correlated to classification accuracy.
- Generative replay CNNs can significantly reduce the need for recalibration by minimizing catastrophic forgetting.

5.2 Practical Implications of Findings

Implications for training data collection

All three aims of this dissertation have findings that are relevant to training data collection protocols. Covariate shift in EMG signals is often caused by sample selection bias. Thus, an obvious way to reduce this type of covariate shift is to collect more varied training samples. However, this process can be cumbersome and would have to be repeated for every recalibration session. In the first aim, we demonstrated that simply moving their residual limb around in their workspace during training data collection can eliminate the effects of limb position in amputees. This training data collection protocol was easy to perform and took the same amount of time as the standard training data collection protocol, demonstrating a convenient way to reduce sample selection bias. In the second aim, we applied a data augmentation protocol that altered signals to resemble examples of interface noise. This augmented data set proved useful, as it increased robustness to interface noise. Notably, the CNN classifier had much poorer results when data augmentation was not used, emphasizing the importance of data augmentation. Thus, this is another method to reduce sample selection bias without needing to collect more data. Finally, in the third aim, the generative replay component of the gr-CNN added more training samples during

recalibration to prevent overfitting. As a result, the CNN classifier could be successfully recalibrated with just one gesture repetition of new training data.

Implications for adaptive LDA algorithms

Although this work was focused on deep learning tools, I compared them to two adaptive LDA algorithms in Chapters 3 and 4. In Chapter 3, the LDA- algorithm removed noisy channels before classification. When I simulated a perfect fault detector, this method yielded the best accuracies across all classification strategies. The simplicity of the algorithm makes it an attractive solution, but the main limitation is that it hinges on the choice of a fault detector. In the original study, this fault detector had error rates of around 30%. With this fault detector, the LDA- algorithm may not be a desirable method. However, if a better fault detector existed, it would be worthwhile to evaluate if the simplicity of the LDA- algorithm would make it a better choice than the CNN classifier. In Chapter 4, the a-LDA calibration technique was the second-best method, behind gr-CNN. The concern with this method is that it had significantly higher proportions of failed recalibrations. Real-time comparisons these methods may be able to tell us if these failed recalibrations are clinically significant, which would inform our decision between the gr-CNN and a-LDA methods. In any case, these strategies proved to be effective and are a viable option for prosthetic systems that may not be able to handle deep learning calculations.

5.3 Limitations and Future Directions

A major limitation of the work in this dissertation is that it lacks online validation. As emphasized by the results in Aim 1, offline results may not be representative of real-time behaviors. Although they are convenient for iterating and developing machine learning techniques, offline analyses exclude crucial factors that influence control performance, such as sensory feedback and user adaptation. Additionally, the metrics and outcome measures of our techniques

were limited mostly to offline accuracy. Ultimately, none of these measures can replace functional evaluations and patient feedback. Thus, an immediate future step is to implement the methods from Chapter 3 and 4 in real time.

On a related note, the deep learning models were all computed on a laptop computer with a GPU. It remains to be seen whether the feasibility of these methods will extend to simpler devices such as the embedded systems used to control prostheses. It is expected that the initial training of the deep convolutional neural network can be performed on a computer and all subsequent adaptations can be performed on-device, which is a benefit of the fast recalibration techniques investigated in Chapter 4.

To maintain clinical feasibility, we used time domain and autoregressive features for all our classification strategies. A major benefit of deep learning models is that they can extract salient features from high-dimensional data. By reducing our inputs to these time domain and autoregressive features, we are not fully taking advantage of the capabilities of deep learning. Previous studies have successfully used convolutional networks with more complex inputs like raw signals and spectrograms. These inputs more closely resemble images or speech data, which are the most common data types used to develop deep learning tools. Thus, these features are likely to obtain better performances. However, our decision to use time domain features was a conscious one to maintain clinical feasibility and see if deep learning models can still be effective with different types of inputs. Nevertheless, it would be beneficial to compare the benefits and drawbacks of different network inputs to determine the optimal network input that balances performance and feasibility.

The control strategies in this work were limited to sequential pattern recognition control. The process of adjusting a neural network to perform regression is straightforward and would be

an appealing next step as simultaneous control is consistently requested by amputee users. The tools developed in this dissertation can all be used with a regression based neural network, including the virtual reality environment, data augmentation protocol, and recalibration techniques.

References

Aljundi, R. (2019). *Continual Learning in Neural Networks*. *September*.

Ameri, A., Akhaee, M. A., Scheme, E., & Englehart, K. (2019). Regression convolutional neural network for improved simultaneous EMG control. *Journal of Neural Engineering*, *16*(3). https://doi.org/10.1088/1741-2552/ab0e2e

Ameri, A., Akhaee, M. A., Scheme, E., Member, S., Englehart, K., & Member, S. (2020). *A Deep Transfer Learning Approach to Reducing the Effect of Electrode Shift in EMG Pattern Recognition-Based Control*. *28*(2), 370–379.

Atzori, M., Cognolato, M., & Müller, H. (2016). Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands. *Frontiers in Neurorobotics*, *10*(SEP), 1–10. https://doi.org/10.3389/fnbot.2016.00009

Bajaj, N. M., Spiers, A. J., & Dollar, A. M. (2015). State of the art in prosthetic wrists: Commercial and research devices. *IEEE International Conference on Rehabilitation Robotics*, *2015-Septe*, 331–338. https://doi.org/10.1109/ICORR.2015.7281221

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7700 LECTU*, 437–478. https://doi.org/10.1007/978-3-642-35289-8_26

Beran, R. (1977). Minimum Hellinger Distance Estimates for Parametric Models. *The Annals of Statistics*, *5*(3), 445–463.

Betthauser, J. L., Hunt, C. L., Osborn, L. E., & Masters, M. R. (2018). *Limb Position Tolerant Pattern Recognition for Myoelectric Prosthesis Control with Adaptive Sparse Representations From Extreme Learning*. *65*(4), 770–778. https://doi.org/10.1109/TBME.2017.2719400

Biddiss, E., & Chau, T. (2007a). Upper-limb prosthetics: Critical factors in device abandonment. *American Journal of Physical Medicine and Rehabilitation*, *86*(12), 977–987. https://doi.org/10.1097/PHM.0b013e3181587f6c

Biddiss, E., & Chau, T. (2007b). Upper limb prosthesis use and abandonment: A survey of the last 25 years. *Prosthetics and Orthotics International*, *31*(3), 236–257. https://doi.org/10.1080/03093640600994581

Boschmann, A., & Platzner, M. (2014). Towards robust HD EMG pattern recognition: Reducing electrode displacement effect using structural similarity. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*, 4547–4550. https://doi.org/10.1109/EMBC.2014.6944635

Campbell, E., Member, S., Cameron, J. A. D., Scheme, E., & Member, S. (2020). *Feasibility of*

Data-driven EMG Signal Generation using a Deep Generative Model. 3755–3758.

Campbell, E., Phinyomark, A., & Scheme, E. (2020). *Current Trends and Confounding Factors in Myoelectric Control : Limb Position and Contraction Intensity*. 1–44. https://doi.org/10.3390/s20061613

Castellini, C., Fiorilla, A. E., & Sandini, G. (2009). Multi-subject/daily-life activity EMG-based control of mechanical hands. *Journal of NeuroEngineering and Rehabilitation*, *6*(1), 1–11. https://doi.org/10.1186/1743-0003-6-41

Chollet, F., & others. (2015). *Keras*.

Chowdhury, R. H., Reaz, M. B. I., Bin Mohd Ali, M. A., Bakar, A. A. A., Chellappan, K., & Chang, T. G. (2013). Surface electromyography signal processing and classification techniques. *Sensors (Switzerland)*, *13*(9), 12431–12466. https://doi.org/10.3390/s130912431

Cipriani, C., Controzzi, M., Kanitz, G., & Sassu, R. (2012). The effects of weight and inertia of the prosthesis on the sensitivity of electromyographic pattern recognition in relax state. *Journal of Prosthetics and Orthotics*, *24*(2), 86–92. https://doi.org/10.1097/JPO.0b013e3182524cce

Coapt LLC. (n.d.). *COMPLETE CONTROL*.

Cordella, F., Ciancio, A. L., Sacchetti, R., Davalli, A., Cutti, A. G., Guglielmelli, E., & Zollo, L. (2016). Literature review on needs of upper limb prosthesis users. *Frontiers in Neuroscience*, *10*(MAY), 1–14. https://doi.org/10.3389/fnins.2016.00209

Côté-Allard, U., Fall, C. L., Drouin, A., Campeau-Lecours, A., Gosselin, C., Glette, K., Laviolette, F., & Gosselin, B. (2019). Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*(4), 760–771. https://doi.org/10.1109/TNSRE.2019.2896269

D., Y., W., Y., Q., H., & H., L. (2017). Classification of Multiple Finger Motions during Dynamic Upper Limb Movements. *IEEE Journal of Biomedical and Health Informatics*, *21*(1), 134–141. https://doi.org/10.1109/JBHI.2015.2490718 LK - http://rug.on.worldcat.org/atoztitles/link/?sid=EMBASE&issn=21682194&id=doi:10.1109%2FJBHI.2015.2490718&atitle=Classification+of+Multiple+Finger+Motions+during+Dynamic+Upper+Limb+Movements&stitle=IEEE+J.+Biomedical+Health+Informat.&title=IEEE+Journal+of+Biomedical+and+Health+Informatics&volume=21&issue=1&spage=134&epage=141&aulast=Yang&aufirst=Dapeng&auinit=D.&aufull=Yang+D.&coden=ITIBF&isbn=&pages=134-141&date=2017&auinit1=D&auinitm=

Datta, D., Selvarajah, K., & Davey, N. (2004). Functional outcome of patients with proximal upper limb deficiency - Acquired and congenital. *Clinical Rehabilitation*, *18*(2), 172–177. https://doi.org/10.1191/0269215504cr716oa

Davidson, J. (2002). A survey of the satisfaction of upper limb amputees with their prostheses, their lifestyles, and their abilities. *Journal of Hand Therapy*, *15*(1), 62–70.

https://doi.org/10.1053/hanthe.2002.v15.01562

De Luca, C. J., Gilmore, L. D., Kuznetsov, M., & Roy, S. H. (2010). Filtering the surface EMG signal movement artifact and baseline noise contamination. *Journal of Biomechanics*, 1573–1579.

Delange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., Slabaugh, G., & Tuytelaars, T. (2021). A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *c*, 1–29. https://doi.org/10.1109/TPAMI.2021.3057446

Dillingham, T. R., Pezzin, L. E., & MacKenzie, E. J. (2002). Limb amputation and limb deficiency: Epidemiology and recent trends in the United States. In *Southern Medical Journal* (Vol. 95, Issue 8, pp. 875–883). https://doi.org/10.1097/00007611-200208000-00018

Engdahl, S. M., Christie, B. P., Kelly, B., Davis, A., Chestek, C. A., & Gates, D. H. (2015). *Surveying the interest of individuals with upper limb loss in novel prosthetic control techniques*. 1–11. https://doi.org/10.1186/s12984-015-0044-2

Englehart, K., & Hudgins, B. (2003). A robust, real-time control scheme for multifunction myoelectric control. *IEEE Transactions on Bio-Medical Engineering*, *50*(7), 848–854. https://doi.org/10.1109/TBME.2003.813539

Fernandez, A., Isusi, I., & Gomez, M. (2000). Factors conditioning the return to work of upper limb amputees in Asturias, Spain. *Prosthetics and Orthotics International*, *24*(2), 143–147. https://doi.org/10.1080/03093640008726537

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6).

Fougner, A, Scheme, E. J., Chan, a D. C., Englehart, K., & Stavdahl, Ø. (2011). A multi-modal approach for hand motion classification using surface EMG and accelerometers. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, *2011*(Grant 192546), 4247–4250. https://doi.org/10.1109/IEMBS.2011.6091054

Fougner, Anders, Scheme, E. J., Chan, A. D. C., Englehart, K., & Stavdahl, Ø. (2011). Resolving the limb position effect in myoelectric pattern recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *19*(6), 644–651. https://doi.org/10.1109/TNSRE.2011.2163529

Fraser, G. D., Chan, A. D. C., Green, J. R., Abser, N., & MacIsaac, D. (2011). CleanEMG - Power line interference estimation in sEMG using an adaptive least squares algorithm. *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *1*, 7941–7944. https://doi.org/10.1109/IEMBS.2011.6091958

Gazzoni, M., Farina, D., & Merletti, R. (2004). A new method for the extraction and classification of single motor unit action potentials from surface EMG signals. *Journal of*

Neuroscience Methods, *136*(2), 165–177. https://doi.org/10.1016/j.jneumeth.2004.01.002

Geng, W., Du, Y., Jin, W., Wei, W., Hu, Y., & Li, J. (2016). Gesture recognition by instantaneous surface EMG images. *Scientific Reports*, *6*(October), 6–13. https://doi.org/10.1038/srep36571

Geng, Y., Zhang, F., Yang, L., Zhang, Y., & Li, G. (2012). Reduction of the effect of arm position variation on real-time performance of motion classification. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2772–2775. https://doi.org/10.1109/EMBC.2012.6346539

Geng, Y., Zhou, P., & Li, G. (2012). Toward attenuating the impact of arm positions on electromyography pattern-recognition based motion classification in transradial amputees. *Journal of NeuroEngineering and Rehabilitation*, *9*(1), 1–11. https://doi.org/10.1186/1743-0003-9-74

Goldenberg, I., & Webb, G. I. (2019). Survey of distance measures for quantifying concept drift and shift in numeric data. *Knowledge and Information Systems*, *60*(2), 591–615. https://doi.org/10.1007/s10115-018-1257-z

Hahne, Janne M., Schweisfurth, M. A., Koppe, M., & Farina, D. (2018). Simultaneous control of multiple functions of bionic hand prostheses: Performance and robustness in end users. *Science Robotics*, *3*(19), eaat3630. https://doi.org/10.1126/scirobotics.aat3630

Hahne, Janne Mathias, Graimann, B., & Muller, K. R. (2012). Spatial filtering for robust myoelectric control. *IEEE Transactions on Biomedical Engineering*, *59*(5), 1436–1443. https://doi.org/10.1109/TBME.2012.2188799

Hargrove, L. J., Lock, B. A., & Simon, A. M. (2013). Pattern recognition control outperforms conventional myoelectric control in upper limb patients with targeted muscle reinnervation. *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1599–1602. https://doi.org/10.1109/EMBC.2013.6609821

Hargrove, L., Losier, Y., Lock, B., Englehart, K., & Hudgins, B. (2007). A real-time pattern recognition based myoelectric control usability study implemented in a virtual environment. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 4842–4845. https://doi.org/10.1109/IEMBS.2007.4353424

Hargrove, Levi, Miller, L., Turner, K., & Kuiken, T. (2018). Control within a virtual environment is correlated to functional outcomes when using a physical prosthesis. *Journal of NeuroEngineering and Rehabilitation*, *15*(Suppl 1). https://doi.org/10.1186/s12984-018-0402-y

Holden, D., Saito, J., Komura, T., & Joyce, T. (2015). Learning motion manifolds with convolutional autoencoders. *SIGGRAPH Asia 2015 Technical Briefs, SA 2015*, 1–4. https://doi.org/10.1145/2820903.2820918

Hu, B., Simon, A. M., & Hargrove, L. (2019). Deep generative models with data augmentation to learn robust representations of movement intention for powered leg prostheses. *IEEE*

Transactions on Medical Robotics and Bionics, *1*(4), 267–278. https://doi.org/10.1109/tmrb.2019.2952148

Hudgins, B., Parker, P., & Scott, R. N. (1993). A new strategy for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, *40*(1), 5541–5548. https://doi.org/10.1109/10.204774

Hwang, H. J., Hahne, J. M., & Müller, K. R. (2017). Real-time robustness evaluation of regression based myoelectric control against arm position change and donning/doffing. *PLoS ONE*, *12*(11), 1–22. https://doi.org/10.1371/journal.pone.0186318

Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Proceedings of the 32nd International Conference on Machine Learning*, *37*. https://doi.org/10.1080/17512786.2015.1058180

Ison, M., & Artemiadis, P. (2014). The role of muscle synergies in myoelectric control: Trends and challenges for simultaneous multifunction control. *Journal of Neural Engineering*, *11*(5). https://doi.org/10.1088/1741-2560/11/5/051001

Jang, C. H., Yang, H. S., Yang, H. E., Lee, S. Y., Kwon, J. W., Yun, B. D., Choi, J. Y., Kim, S. N., & Jeong, H. W. (2011). A Survey on Activities of Daily Living and Occupations of Upper Extremity Amputees. *Annals of Rehabilitation Medicine*, *35*(6), 907. https://doi.org/10.5535/arm.2011.35.6.907

Jiang, N., Englehart, K., & Parker, P. (2009). Extracting simultaneous and proportional neural control information for multiple-DOF prostheses from the surface electromyographic signal. *IEEE Transactions on Biomedical Engineering*, *56*(4), 1070–1080. https://doi.org/10.1109/TBME.2008.2007967

Jiang, Ning, Rehbaum, H., Vujaklija, I., Graimann, B., & Farina, D. (2013). Intuitive, Online, Simultaneous and Proportional Myoelectric Control Over Two Degrees of Freedom in Upper Limb Amputees. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *22*(3), 501–510. https://doi.org/10.1109/TNSRE.2013.2278411

Kemker, R., Mcclure, M., Abitino, A., Hayes, T. L., & Kanan, C. (2017). *Measuring Catastrophic Forgetting in Neural Networks*. 3390–3398.

Kendall, F. P., & McCreary, E. K. (1983). *Muscles Testing and Function* (3rd ed.). Williams & Wilkens.

Khushaba, R. N., Takruri, M., Miro, J. V., & Kodagoda, S. (2014). Towards limb position invariant myoelectric pattern recognition using time-dependent spectral features. *Neural Networks*, *55*, 42–58. https://doi.org/10.1016/j.neunet.2014.03.010

Kingma, D. P., & Ba, J. L. (2015). ADAM: A Method for Stochastic Optimization. *Proceedings of the 3rd International Conference on Learning Representations*. https://arxiv.org/pdf/1412.6980.pdf %22 entire document

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep

Convolutional Neural Networks. *Advances in Neural Information Processing Systems*. https://doi.org/10.1201/9781420010749

Kyranou, I., Vijayakumar, S., & Erden, M. S. (2018). Causes of performance degradation in non-invasive electromyographic pattern recognition in upper limb prostheses. *Frontiers in Neurorobotics*, *12*(September), 1–22. https://doi.org/10.3389/fnbot.2018.00058

L. Chen, Y. Geng, and G. L. (2011). Effect of upper-limb positions on mo- tion pattern recognition using electromyography. *Conf Image Signal Process (CISP)*, 139–142.

Lecun, Y. (1987). *Modeles connexionistes de l'apprentissage*. Universit´e de Paris VI.

Lecun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time-Series. *The Handbook of Brain Theory and Neural Networks*, *3361*(10), 63–66. https://doi.org/10.1177/016555157900100111

Liu, J., Zhang, D., Sheng, X., & Zhu, X. (2014). Quantification and solutions of arm movements effect on sEMG pattern recognition. *Biomedical Signal Processing and Control*, *13*(1), 189–197. https://doi.org/10.1016/j.bspc.2014.05.001

López, N. M., di Sciascio, F., Soria, C. M., & Valentinuzzi, M. E. (2009). Robust EMG sensing system based on data fusion for myoelectric control of a robotic arm. *BioMedical Engineering Online*, *8*, 1–13. https://doi.org/10.1186/1475-925X-8-5

Luo, T., Zhang, X., We, L., Chen, X., Chen, X., & Chen, X. (2019). Convolutional neural network with data augmentation for robust myoelectric control. *IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*.

Maier, J., Naber, A., & Ortiz-Catalan, M. (2018). Improved prosthetic control based on myoelectric pattern recognition via wavelet-based de-noising. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(2), 506–514. https://doi.org/10.1109/TNSRE.2017.2771273

Miller, L. A., Turner, K., & Simon, A. M. (2020). Data logging during pattern recognition calibration as a remote diagnostic tool. *Myoelectric Controls Symposium*.

Montagnani, F., Controzzi, M., & Cipriani, C. (2015). Is it Finger or Wrist Dexterity That is Missing in Current Hand Prostheses? *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *23*(4), 600–609. https://doi.org/10.1109/TNSRE.2015.2398112

Moreno-Torres, J. G., Raeder, T., Alaiz-Rodriguez, R., Chawla, N. V., & Herrera, F. (2012). A unifying view on dataset shift in classification. *Pattern Recognition*.

Muceli, S., Jiang, N., & Farina, D. (2014). Extracting signals robust to electrode number and shift for online simultaneous and proportional myoelectric control by factorization algorithms. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *22*(3), 623–633. https://doi.org/10.1109/TNSRE.2013.2282898

Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines Vinod. *Proceedings of the 27th International Conference on Machine Learning*. https://doi.org/10.1123/jab.2016-0355

Ng, A. Y. (2004). Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proceedings of the 21st International Conference on Machine Learning*.

Nilsson, N., Håkansson, B., & Ortiz-Catalan, M. (2017). Classification complexity in myoelectric pattern recognition. *Journal of NeuroEngineering and Rehabilitation*, *14*(1), 1–18. https://doi.org/10.1186/s12984-017-0283-5

Ortiz-Catalan, M., Rouhani, F., Branemark, R., & Hakansson, B. (2015). Offline accuracy: A potentially misleading metric in myoelectric pattern recognition for prosthetic control. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2015-Novem*, 1140–1143. https://doi.org/10.1109/EMBC.2015.7318567

Ortolan, R. L., Mori, R. N., Pereira, R. R., Cabral, C. M. N., Pereira, J. C., & Cliquet, A. (2003). Evaluation of adaptive/nonadaptive filtering and wavelet transform techniques for noise reduction in EMG mobile acquisition equipment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *11*(1), 60–69. https://doi.org/10.1109/TNSRE.2003.810432

Ottobock. (n.d.). *Myo Plus TR pattern recognition Private Payer Reimbursement Guide*.

Pan, L., Harmody, A., & Huang, H. (2018). A Reliable Multi-User EMG Interface Based on A Generic-Musculoskeletal Model against Loading Weight Changes ∗. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2018-July*, 2104–2107. https://doi.org/10.1109/EMBC.2018.8512685

Park, K. H., & Lee, S. W. (2016). Movement intention decoding based on deep learning for multiuser myoelectric interfaces. *4th International Winter Conference on Brain-Computer Interface, BCI 2016*, 7–8. https://doi.org/10.1109/IWW-BCI.2016.7457459

Parker, P., Englehart, K., & Hudgins, B. (2006). Myoelectric signal processing for control of powered limb prostheses. *Journal of Electromyography and Kinesiology*, *16*(6), 541–548. https://doi.org/10.1016/j.jelekin.2006.08.006

Pellegrini, L., Graffieti, G., Lomonaco, V., & Maltoni, D. (2020). Latent replay for real-time continual learning. *IEEE International Conference on Intelligent Robots and Systems*, 10203–10209. https://doi.org/10.1109/IROS45743.2020.9341460

Phinyomark, A., Campbell, E., & Scheme, E. (2020). Surface electromyography (EMG) signal processing, classification, and practical considerations. In *Biomedical Signal Processing* (pp. 3–29). Springer Singapore. https://doi.org/10.1007/978-981-13-9097-5_1

Phinyomark, A., Limsakul, C., & Phukpattaranont, P. (2009). A comparative study of wavelet denoising for multifunction myoelectric control. *International Conference on Computer and Automation Engineering*, 21–25. https://doi.org/10.1109/ICCAE.2009.57

Powar, O. S., Chemmangat, K., & Figarado, S. (2018). A novel pre-processing procedure for enhanced feature extraction and characterization of electromyogram signals. *Biomedical Signal Processing and Control*, *42*, 277–286.

Radmand, A., Scheme, E. J., & Englehart, K. (2014). A characterization of the effect of limb position on EMG features to guide the development of effective prosthetic control schemes. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2014*, 662–667. https://doi.org/10.1109/EMBC.2014.6943678

Radmand, Ashkan, Scheme, E. J., & Englehart, K. (2014). On the suitability of integrating accelerometry data with electromyography signals for resolving the effect of changes in limb position during dynamic limb movement. *Journal of Prosthetics and Orthotics*, *26*(4), 185–193. https://doi.org/10.1097/JPO.0000000000000041

Reaz, M. B. I., Hussain, M. S., & Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: Detection, processing, classification and applications. *Biological Procedures Online*, *8*(1), 11–35. https://doi.org/10.1251/bpo115

Rehbaum, H., & Farina, D. (2015). Adaptive common average filtering for myocontrol applications. *Medical and Biological Engineering and Computing*, *53*(2), 179–186. https://doi.org/10.1007/s11517-014-1215-1

Rolnick, D., Schwarz, J., Lillicrap, T. P., & Wayne, G. (2019). *Experience Replay for Continual Learning*. *NeurIPS*.

Routhier, F., Vincent, C., Morissette, M. J., & Desaulniers, L. (2001). Clinical results of an investigation of paediatric upper limb myoelectric prosthesis fitting at the Quebec Rehabilitation Institute. *Prosthetics and Orthotics International*, *25*(2), 119–131. https://doi.org/10.1080/03093640108726585

Salminger, S., Stino, H., Pichler, L. H., Gstoettner, C., Mayer, J. A., Szivak, M., Aszmann, O. C., Salminger, S., Stino, H., Pichler, L. H., Gstoettner, C., Sturma, A., Mayer, J. A., Szivak, M., & Aszmann, O. C. (2020). Current rates of prosthetic usage in upper-limb amputees – have innovations had an impact on device acceptance ? *Disability and Rehabilitation*, *0*(0), 1–12. https://doi.org/10.1080/09638288.2020.1866684

Scheme, E., Biron, K., & Englehart, K. (2011). Improving myoelectric pattern recognition positional robustness using advanced training protocols. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 4828–4831. https://doi.org/10.1109/IEMBS.2011.6091196

Scheme, E. J., & Englehart, K. (2011). Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use. *The Journal of Rehabilitation Research and Development*, *48*(6), 643–660. https://doi.org/10.1682/JRRD.2010.09.0177

Scheme, E. J., Fougner, A., Stavdahl, Chan, A. D. C., & Englehart, K. (2010). Examining the adverse effects of limb position on pattern recognition based myoelectric control. *2010*

Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10, 6337–6340. https://doi.org/10.1109/IEMBS.2010.5627638

Scheme, E. J., Lock, B., Hargrove, L., Hill, W., Kuruganti, U., & Englehart, K. (2014). Motion normalized proportional control for improved pattern recognition-based myoelectric control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *22*(1), 149–157. https://doi.org/10.1109/TNSRE.2013.2247421

Sensinger, J. W., Lock, B. A., & Kuiken, T. A. (2009). Adaptive pattern recognition of myoelectric signals: Exploration of conceptual framework and practical algorithms. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *17*(3), 270–278. https://doi.org/10.1109/TNSRE.2009.2023282

Shin, H., Lee, J. K., Kim, J., & Kim, J. (2013). Continual learning with deep generative replay. *2013 21st Signal Processing and Communications Applications Conference, SIU 2013*, *Nips*. https://doi.org/10.1109/SIU.2013.6531162

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0197-0

Simon, A. M., Lock, B. A., & Stubblefield, K. A. (2012). Patient training for functional use of pattern recognition-controlled prostheses. *Journal of Prosthetics and Orthotics*, *24*(2), 56–64. https://doi.org/10.1097/JPO.0b013e3182515437.Patient

Simon, A. M., Stern, K., & Hargrove, L. J. (2011). A comparison of proportional control methods for pattern recognition control. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 3354–3357. https://doi.org/10.1109/IEMBS.2011.6090909

Simon, A. M., Turner, K. L., Miller, L. A., Hargrove, L. J., & Kuiken, T. A. (2019). Pattern recognition and direct control home use of a multi-articulating hand prosthesis. *IEEE International Conference on Rehabilitation Robotics*, 386–391. https://doi.org/10.1109/ICORR.2019.8779539

Stachaczyk, M., Farokh Atashzar, S., & Farina, D. (2020). Adaptive spatial filtering of high-density EMG for reducing the influence of noise and artefacts in Mmyoelectric control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(7), 1511–1517. https://doi.org/10.1109/TNSRE.2020.2986099

Teh, Y., & Hargrove, L. J. (2020). Understanding limb position and external load effects on real-time pattern recognition control in amputees. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(7), 1605–1613. https://doi.org/10.1109/TNSRE.2020.2991643

Teh, Y., & Hargrove, L. J. (2021). Using latent representations of muscle activation patterns to mitigate myoelectric interface noise. *10th International IEEE/EMBS Conference on Neural Engineering*, 1148–1151. https://doi.org/10.1109/NER49283.2021.9441396

Tommasi, T., Orabona, F., Castellini, C., & Caputo, B. (2013). Improving control of dexterous

hand prostheses using adaptive learning. *IEEE Transactions on Robotics*, *29*(1), 207–219. https://doi.org/10.1109/TRO.2012.2226386

ur Rehman, M. Z., Gilani, S. O., Waris, A., Niazi, I. K., Slabaugh, G., Farina, D., & Kamavuako, E. N. (2018). Stacked sparse autoencoders for EMG-based classification of hand motions: A comparative multi day analyses between surface and intramuscular EMG. *Applied Sciences (Switzerland)*, *8*(7). https://doi.org/10.3390/app8071126

van de Ven, G. M., & Tolias, A. S. (2019). *Three scenarios for continual learning*. 1–18. http://arxiv.org/abs/1904.07734

Ven, G. M. Van De, & Tolias, A. S. (2019). *Generative replay with feedback connections as a general strategy for continual learning*. *April*, 1–17.

Wang, Q., Qin, Z., Nie, F., & Yuan, Y. (n.d.). *Convolutional 2D LDA for Nonlinear Dimensionality Reduction ∗*. 2929–2935.

Webb, G. I., Hyde, R., Cao, H., Nguyen, H. L., & Petitjean, F. (2016). Characterizing concept drift. *Data Mining and Knowledge Discovery*, *30*(4), 964–994. https://doi.org/10.1007/s10618-015-0448-4

Whitney, D. E. (1969). Resolved Motion Rate Control of Manipulators and Human Prostheses. *IEEE Transactions on Man-Machine Systems*, *10*(2), 47–53. https://doi.org/10.1109/TMMS.1969.299896

Williams, T. W., Meier, R. H., & Atkins, D. (2004). Control of powered upper extremity prostheses. *Functional Restoration of Adults and Children with Upper Extremity Amputation*, 207–224.

Woodward, R. B., & Hargrove, L. J. (2019). Adapting myoelectric control in real-time using a virtual environment. *Journal of NeuroEngineering and Rehabilitation*, *16*(1), 1–12. https://doi.org/10.1186/s12984-019-0480-5

Wu, L., Zhang, X., Wang, K., Chen, X., & Chen, X. (2020). Improved High-density Myoelectric Pattern Recognition Control Against Electrode Shift Using Data Augmentation and Dilated Convolutional Neural Network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *4320*(c), 1–1. https://doi.org/10.1109/tnsre.2020.3030931

Yang, D., Gu, Y., Jiang, L., Osborn, L., & Liu, H. (2017). Dynamic training protocol improves the robustness of PR-based myoelectric control. *Biomedical Signal Processing and Control*, *31*, 249–256. https://doi.org/10.1016/j.bspc.2016.08.017

Yang, D., Liu, H., & Member, S. (2021). An EMG-based deep learning approach for multi-DOF wrist movement decoding. *IEEE Transactions on Industrial Electronics*, *0046*(52075114). https://doi.org/10.1109/TIE.2021.3097666

Yang, W., Yang, D., Liu, Y., & Liu, H. (2019). Decoding simultaneous multi-DOF wrist movements from raw EMG signals using a convolutional neural network. *IEEE Transactions on Human-Machine Systems*, *49*(5), 411–420.

https://doi.org/10.1109/THMS.2019.2925191

Young, A. J., Hargrove, L. J., & Kuiken, T. A. (2011). The effects of electrode size and orientation on the sensitivity of myoelectric pattern recognition systems to electrode shift. *IEEE Transactions on Biomedical Engineering*, *58*(9), 2537–2544. https://doi.org/10.1109/TBME.2011.2159216

Zhang, X., & Huang, H. (2015). A real-time, practical sensor fault-tolerant module for robust EMG pattern recognition. *Journal of NeuroEngineering and Rehabilitation*, *12*(1), 1–16. https://doi.org/10.1186/s12984-015-0011-y

Ziegler-Graham, K., MacKenzie, E. J., Ephraim, P. L., Travison, T. G., & Brookmeyer, R. (2008). Estimating the Prevalence of Limb Loss in the United States: 2005 to 2050. *Archives of Physical Medicine and Rehabilitation*, *89*(3), 422–429. https://doi.org/10.1016/j.apmr.2007.11.005